Discovery and Decision: Exploring the Metaphysics and Epistemology of Scientific Classification

Rebecca Bryant

Ph.D.
The University of Edinburgh
1997

Contents

Declaration.

I declare that this thesis has been composed by myself and that the research reported here has been conducted by myself unless otherwise indicated.

Rebecca Bryant

Edinburgh, January 15, 1997.

Acknowledgements.

At the top of my thank you list is my supervisor, Peter Lewis. He has given me endless support by providing me with numerous references, reading reams of my written work and giving up valuable time for really helpful and interesting discussions. He also gave me the idea of using case studies in my work, without which this thesis would have been a very different and, I think, inferior animal.

Thanks also to Alexander Bird for helpful comments and discussion about Chapter Seven and for taking the trouble to read an earlier draft of this thesis in its entirety.

On the less work-oriented front, there are a whole host of people to thank. Helen Wilson, Isla Jack, Sandy Boyle, Alexandra Duer, Darren (Alex) Moseley, Benedict Young, Alice Drewery, Kerstin Mertens and Majeda Omar are some of the people who have been sources of support and encouragement in all sorts of different ways. Hild Leslie deserves an extra special mention here – for endless coffees and chats and above all for sharing my belief that there *are* other ways to think.

Most importantly, I am indebted to my partner, Steffan. Without his continuous love, support and understanding, I probably would never have embarked upon this thesis, let alone finished it.

Abstract.

This thesis comprises an interdisciplinary examination of the phenomenon of scientific classification. The major aims of the thesis are two-fold:

- To show that scientific classification comprises a mixture of metaphysics and epistemology.
- To illustrate that the psychological explanation-based account of categorisation is an appropriate model for scientific as well as for lay classification.

I begin by characterising the doctrine of Objectivism which provides the philosophical backdrop for traditional accounts of categorisation in both philosophy and psychology. I illustrate some of the shortcomings of Objectivism and propose an alternative doctrine in its place – internal or experiential realism. This alternative doctrine forms the philosophical backdrop for my own account of (scientific) classification.

Next, I provide a short history of theories in the psychology of categorisation. I argue that there has been a shift from very tightly defined, inflexible and context-insensitive accounts (the classical view) to much more flexible, context-sensitive and human-oriented accounts (the explanation-based view). I go on to to examine the criticisms which an objectivist philosopher, Georges Rey, has levied against work in the psychology of categorisation.

I then put forward my own positive account of scientific classification. This incorporates a number of claims: that classification involves more than a reflection of metaphysics, that human beings (and so epistemology) make an active contribution to the classificatory process, that the dividing line between metaphysics and epistemology is blurred, that there are problems with applying the objectivist doctrine of essentialism to scientific classification and as a result of the foregoing points, that the explanation-based account of categorisation is a suitable model of classification by scientific experts. The material which I use in support of these

claims comes from three case studies in the history and sociology of science (which are supported by a review of the debate concerning species concepts in contemporary biology) and from a number of philosophical arguments.

Finally, I give a more theoretical account of why the explanation-based view is a suitable model of scientific classification. I also argue that there are ways in which the scientist is constrained in classifying entities in the natural world and that these constraints are linked by the mechanism of scientific explanation. I point out, however, that the doctrine of Objectivism cannot be re-introduced at the level of scientific explanation.

"It is now evident that where one discipline ends and the other begins no longer matters, for it is in the nature of the case that the boundaries are ill-defined."

(Patricia Churchland, *Neurophilosophy*, p.x.)

"We find one truth and embrace it. Then we close our eyes to everything else. It avoids confusion."

(David Eddings, *The Elenium*, p.164.)

Chapter 1
Introduction.

Aims of Thesis.

There is a philosophical tradition dating back to Aristotle which claims that the natural world possesses a unique structure in and of itself. In particular, this tradition claims that the world is divided inherently into classes and that the members of each class share a common essence or nature which takes the form of a set of defining characteristics. The essence comprises *what it is* to be a member of the relevant class or kind. This doctrine of essentialism of kinds[1] has implications for both classification and definition. It is argued that to classify entities correctly is simply to reproduce the world's inherent structure. To give the correct definition of what it is to be a particular kind of thing and so to provide the meaning of the word which refers to that kind of thing is to reiterate the essence or the defining characteristics of the natural class to which it belongs.

The doctrine of essentialism entails a number of cognate assumptions. These include the notion that correct classification is a matter of metaphysics alone, that epistemology is irrelevant for matters of classification and that human beings play an entirely passive role in the classificatory process by simply *reflecting* the metaphysics.

More recently, this doctrine has been adopted in the early work of Hilary Putnam and by Saul Kripke. These philosophers have linked essentialism with *science*, claiming that it is the role of science to uncover the essential properties which underlie and explain the natural world's inherent division into kinds.

In the context of the position sketched above, the aims of this thesis are two-fold:

- To illustrate that *scientific* classification comprises a *combination of metaphysics and epistemology*.

[1] There is also a doctrine of essentialism of particulars. See, for instance, Kripke (1972). However, since I am concerned with kinds, this doctrine will be ignored.

- In view of this combination, to show that the *psychological explanation-based account of categorisation is a suitable model of expert scientific as well as of lay classification.*

I concentrate on scientific classification partly because of the weight which Putnam and Kripke lend to science in explicating their doctrines and partly because I think it is natural to expect that if the doctrine of essentialism and its cognate assumptions were correct, they would be proved to be so in the scientific domain, if anywhere, simply because this domain is popularly construed as being objective and context-free.

It is important to stress that the position which I argue for in this thesis *is not* correctly described as relativist, subjectivist or anti-realist. Rather, it represents a middle road between the extremes of metaphysical realism and relativism. I *do not* argue that there is no external world nor that human beings somehow construct the world which they inhabit. Rather, my arguments demand the existence of a mind-independent world which is populated by real entities and properties. This position is therefore correctly described as a realist one.

Organisation of Thesis.

Chapter Two comprises an introduction to general theoretical positions which can be seen to motivate differing theories of and approaches to categorisation. I summarise the doctrine of Objectivism and show how this is related to traditional interpretations of categorisation. I then offer Putnam's internal realism and Johnson's experiential realism as more satisfactory alternatives to Objectivism. In Chapter Three, I examine some of the objections levied against internal/experiential realist positions. I also discuss how the Putnam/Johnson approach serves to contextualise my own account of scientific classification and the attention which I pay to empirical evidence.

Chapters Four and Five constitute a characterisation and analysis of the history of the major *psychological* views of classification up to the present day. In particular, in Chapter Four, I suggest that the progress from the early to the later views of categorisation marks a shift from the interpretation of categorisation as a tightly defined process in which humans play an entirely passive role to that of a much more flexible, context-sensitive process in which humans play an instrumental role. Chapter Five considers the criticisms which an objectivist philosopher, Georges Rey,

levies against work in the psychology of categorisation and discusses the responses of the psychologists to these criticisms, particularly in relation to the division which Rey insists must be drawn between metaphysics and epistemology.

The remaining chapters of the thesis present my own positive account of scientific classification. Chapter Six considers three scientific case studies which, I argue, illustrate my claims that scientific classification comprises a mixture of metaphysics and epistemology and that the psychological explanation-based account succeeds as a model of scientific as well as of lay classification. These case studies are supported by material from contemporary biology which illustrates that the debate concerning species concepts is not simply a phenomenon of the history of science and is raging as fiercely today as it ever was. Chapter Seven builds on the conclusions reached in Chapter Six; I summarise papers by the philosophers Keith Donnellan and John Canfield which criticise the Putnam/Kripke approach to natural kinds and I illustrate how their claims can be analysed and extended to lend support to my own account of scientific classification. In Chapter Eight, I give a more theoretical formulation of my own account, indicating how it relates to Putnam's hypothesis of the division of linguistic labour and discussing some associated issues such as ways in which scientists are constrained in their classifications and the connection between classification and explanation.

Finally, Chapter Nine draws all the preceding chapters together and sets out what has been achieved in this thesis.

Methodology.

As can be seen from the preceding section, I utilise a rather unusual methodology for philosophical research. This is so in two ways. Firstly, I take heed of psychological evidence relating to the process of categorisation and secondly, I use case studies from the history and sociology of science coupled with material from contemporary biology in order to support my clams about scientific classification.

This methodology is really a consequence of the stance which I take on the ways in which human beings are situated in, affect and are affected by their environment. This stance can be equated theoretically with Putnam's internal realism and Johnson's experiential realism, as described in Chapter Two of this thesis.

As a result of the methodological framework which I adopt, this thesis is correctly

described as an *interdisciplinary* study of the phenomenon of scientific classification.

Some Terminology.

To begin with, it will be helpful to define some of the basic terms which are used extensively throughout this thesis.

Classification or *categorisation*: this is the process by which human beings group together particular entities and treat them as equivalent in some sense or senses. These two terms are used interchangeably throughout this thesis. For example, we might say that human beings classify particular animals together and treat them as equivalent in that they all have wings, feathers and beaks.

Class, *category* or *kind*: this is a collection of entities which have been grouped together by the process of classification or categorisation. These three terms are used interchangeably throughout this thesis. For example, the collection of animals all of which have wings, feathers and beaks is known as the 'bird' class or category.

Concept: this is a mental representation or mental counterpart of a class, category or kind. For example, the mental representation or idea that human beings have of birds will correspond to the 'bird' class or category.

Using Putnam.

Much of the material which appears in this thesis is a discussion of various works of Hilary Putnam. To avoid confusion, it is useful to note that the material I discuss comes from both his early strong realist period and his later internal realist period. His early period comprises all work prior to the book, *Reason, Truth and History*. His later period comprises *Reason, Truth and History* and all work following this book. I criticise aspects of the early work,[2] whilst using the later work as a backdrop for my positive account of scientific classification.

[2] I concentrate on his paper, "The Meaning of "Meaning"".

Chapter 2
Objectivism and Internal Realism.

> "...we discovered that certain assumptions of contemporary philosophy and linguistics that have been taken for granted within the Western tradition since the Greeks precluded us from even raising the kind of issues we wanted to address. The problem was not one of extending or patching up some existing theory of meaning but of revising central assumptions in the Western philosophical tradition."[3]

Introduction.

In this chapter, I aim to provide a backdrop against which my final account of categorisation makes sense. In order to appreciate this account of categorisation, it is necessary to reject a complete philosophical mind-set – that of Objectivism – and adopt an alternative basic way of thinking about metaphysics and epistemology in which human beings are accorded an important and active role.

In order to do this, I provide a complete account of Objectivism. I then consider the shortcomings of this doctrine and describe an alternative doctrine (internal realism) which has been put forward by Hilary Putnam and further developed by Mark Johnson in which emphasis is placed upon human beings as organisms which constitute a part of and are in continuous interaction with the environment.

Characterising Objectivism.

It is helpful to begin with an overview of what the objectivist picture involves.[4] The most basic objectivist tenet involves a commitment to the existence of a real world which is independent of and external to the human mind and so its experience or cognition. It is assumed that this independent world consists of entities, properties belonging to those entities and relations holding between those entities. Furthermore, it is assumed that the world is *uniquely* structured in terms of these entities,

[3] Lakoff and Johnson, 1980, pp.ix-x.
[4] This characterisation of Objectivism draws heavily upon Lakoff's (1987) account.

properties and relations. In other words, it is assumed that there is *one and only one correct description of reality* – the one that mirrors the structure which is considered to be inherent in the world. As George Lakoff puts it, "...reality is correctly and completely structured in a way that can be modelled by set-theoretical models – that is, in terms of entities, properties and relations. This structure exists, independent of any human understanding."[5]

As far as the objectivist is concerned, words and mental representations are simply abstract symbols which are, in themselves, totally meaningless. When we think or reason, the objectivist sees us as algorithmically manipulating these abstract symbols. The way in which the symbols become meaningful, according to the objectivist, is by being made to correspond to entities and categories in the real or other possible worlds. Hence if we utilise words or mental representations in such a way that they fail to correspond to the world's inherent structure, then they remain meaningless.[6]

It can thus be seen that Objectivism covers two domains – on the one hand, metaphysics and on the other, cognition and language (epistemology). I will now turn to look at these two domains in more detail.

Objectivist Metaphysics.

As stated above, the primary commitment of objectivist metaphysics is to the idea that reality consists of entities with fixed properties and relations holding among them. There is often a further assumption which goes along with this – that some of the properties which these entities have are essential. That is, possession of these essential properties is what makes the entity the thing that it is – the properties capture the essence of what it is to be that thing. This is known as the doctrine of essentialism. Other properties are said to be accidental – they do not capture the essence of what it is to be that thing. For example, we might say that having the property of being a rational featherless biped is what captures the essential nature of human beings. Other properties such as height and skin colour are merely accidental, non-essential properties. They do not contribute to the essential nature of human beings.

---

[5] Lakoff, 1987, p.159.

[6] It is worth noting that false descriptions present a problem for the objectivist, since they are meaningful, yet fail to correspond to the world's inherent structure.

Objectivism makes the further claim that all the entities in the world fall into certain categories or groups. This is known as the classical view of categorisation and it states that all entities in the world which share the same given property or set of properties form a category. These common properties in fact define the category and provide necessary and sufficient conditions for membership in the category. Each essential property is seen as singly necessary and the totality of properties are seen as jointly sufficient for membership in the category. For example, being rational, featherless and a biped might form the necessary and sufficient conditions for being a human being and in order for an entity to belong to the category of human beings, it must possess all those properties.

According to Objectivism, both entities and their properties really and objectively exist. Since the classical view of categorisation is based on the (non)possession of these objective properties by these objective entities, the further claim is made that the categories themselves exist objectively. In other words, not only does the world consist of entities, but these entities all fall into categories which also form part of the structure of the objective world.

In some cases, Objectivism combines the classical view of categorisation with the doctrine of essentialism and thereby comes up with a special type of category known as a natural kind. The notion here is that certain entities in the world fall into certain natural kinds, each kind being a classical category. The basis for membership in a natural kind is possession of certain essential, rather than accidental properties. Thus, the essence which is attached to the natural kind is not a matter of the way language works, but is a matter of the inherent nature and structure of the world. The most commonly cited examples of natural kinds are things which we consider to be natural phenomena, such as animals, plants, chemical elements and compounds.

Objectivism asserts that there are logical relations which objectively exist between the categories of the world. For instance, certain categories will be included within certain other categories and this means that each included category will exhibit the properties exhibited by the category in which it is included. We might say that the category of human beings falls within the category of mammals. If the category of mammals is defined as 'warm-blooded creatures which suckle their young', then all human beings must be warm-blooded and suckle their young. Alternatively, two categories might be mutually exclusive, which means that no member of the one category can be a member of the other category.

Finally, Objectivism asserts what is known as real-world atomism. This doctrine claims that there are certain properties in the world which are basic (atomic) and so cannot be broken down or analysed any further – they have no internal structure. These properties can be combined logically into more complex properties which do have internal structure. The property of being a human being might be a complex property which comprises a combination of the atomic properties of being rational, being featherless and being a biped.

What we can see from this account of objectivist metaphysics is that the objectivist world is meant to be self-contained and to have considerable structure in itself. It consists of fixed entities, properties and relations, these entities falling into fixed categories on the basis of possession of shared properties, with certain logical relations holding between categories. What it is important to realise is that this internal structure is something which the world is believed to have objectively, in itself and *quite independent of human thought, experience or cognition.*

Objectivist Epistemology.

Given objectivist metaphysics, the objectivist has to be able to answer questions about what meaning is, what makes reasoning correct and how it is that we can come to know something about the world. The objectivist key to all these questions is the idea that the mind can *mirror* the structure inherent in the world. In order to gain knowledge or reason correctly, we must be able to reproduce the world and its structure as it is in itself, quite independent of human cognition. Reproducing a correct reflection of the world ensures knowledge of that world.

As stated above, the basis of this approach is that words and mental representations (or concepts) are seen as abstract symbols which only acquire meaning by successfully referring to elements of the objective world. When we reason correctly, we manipulate these symbols according to logical rules so that they reflect the structure of the world and we thereby gain knowledge of the world. In particular, these symbols must correspond to objective entities and categories and we must be able to reproduce the logical relations which are said to exist between these entities and categories. In other words, correct reasoning and knowledge involve *reflecting the inherent and logical structure of the world.*

This kind of approach effectively moves rationality and knowledge away from the personal human sphere, assuming the plausibility of what is known as a God's Eye

perspective – an impersonal, universal, context-free perspective. Ultimate (objective) knowledge and rationality are thus considered to be quite independent of individual human thought, belief or cognition. If we manage to mirror the world's entities, categories and the relations holding between them, then we will have knowledge. If, however, we fail to do this, we will be denied the God's Eye perspective and knowledge will evade us.

So how do human concepts fit into this kind of account of human cognition? Firstly, concepts, for the objectivist, are simply mental representations of the categories and entities of the world. By definition, they must exclude anything which is dependent on human conceptualisation, thought or belief about that world. As Lakoff puts it, "It simply requires that our system of concepts be defined independent of cognitive processing. Aspects of cognition can have nothing whatever to do with truth or meaning or correct reason or real knowledge."[7] Thus the classical view of categorisation covers conceptual categories as well as real-world categories. Conceptual categories will be symbolic representations of categories in the real or other possible worlds. The members of conceptual categories will be symbolic representations of entities in the real or other possible worlds. Conceptual categories will mirror the structure of real-world categories, so they will be defined in terms of necessary and sufficient conditions which will be based on the (non)possession of properties by the symbolic representations of real-world entities. These properties and conditions will of course mirror the properties and conditions which hold in the real or other possible worlds. Conceptual categories simply are mirror images of real-world categories.

The objectivist account of language is very similar to its account of human thought and cognition. Statements are said to be meaningful only if they can be evaluated as true or false. Linguistic expressions (just like mental representations or concepts) acquire meaning only by corresponding to the real or some possible world. This either takes the form of correctly referring, in the case of noun phrases or of being true or false, in the case of sentences. The general idea is that referring expressions correspond to entities and predicates correspond to properties. Sentences will be true when their noun phrases and predicates stand in a relation which corresponds to some state-of-affairs in the actual or some possible world.

The objectivist therefore wants to claim that meaning can be explained in terms of

[7] Lakoff, 1987, p.166.

truth and truth conditions. To give the meaning of a particular sentence is to give the conditions under which it would be true or to give its conditions of satisfaction – the state-of-affairs in the world which would have to obtain in order for that sentence to be satisfied. This is, of course, an objectivist conception of truth which involves accurate reflection of the world's inherent logical structure.

What needs to be emphasised in this account of objectivist epistemology is that, as with objectivist metaphysics, 'correct' cognition and language are seen as quite independent of what humans think and do. In order for our cognition and language to be correct, they must simply mirror the world's inherent structure, untainted by any human element whatsoever. Meaning is simply an objective relation between symbols and the world and is quite independent of any human grasp of meaning. This involves a transcendent, God's Eye view of rationality and knowledge – correctly referring to and reflecting the world's structure ensures that we will have knowledge through God's eyes.

Some Objectivist Philosophers.

The above account constitutes a general characterisation of Objectivism. It would not be correct to say that every objectivist philosopher adheres to every element of the doctrine, as I have described it. Rather, it is more the case that the very general principles which can be seen to run through my account of Objectivism – that there exists an outside world, that to gain knowledge we must mirror the structure of that world, that human interests and purposes are irrelevant to a correct description of the way the world is – form a backdrop against or a context within which objectivist philosophers have couched their own theories. Different philosophers then highlight different aspects of Objectivism in the course of formulating their own theories. It might therefore be useful to take a brief look at some such philosophers and their doctrines.

Three Objectivist Philosophers.

Perhaps the earliest objectivist philosopher is Aristotle. Certainly, modern versions of the doctrine of essentialism date back to the teachings of Aristotle and the psychological classical view of categorisation[8] is taken more-or-less directly from his work.

[8] See Chapter Four.

In the *Categories*, individual entities such as particular men and animals are labelled primary substances, whilst the species and genera of primary substances are labelled secondary substances. For Aristotle, definition is a matter of fitting the thing to be defined into a particular classification scheme. This scheme sorts things into various kinds (the genera) which are themselves sorted into various sub-kinds (the species) which are distinguished from one another by a number of characteristic properties (the differentia). For example, we are told that:

> "(...as for secondary substances, the species is predicated of the individual, the genus both of the species and the individual. Similarly, differentia too are predicated both of the species and of the individuals.) And the primary substances admit the definition of the species and of the genera, and the species admits that of the genus; for everything said of what is predicated will be said of the subject also. Similarly, both the species and the individuals admit the definition of the differentiae."[9]

Thus to give the correct definition of what it is to be a human, for instance, is to provide the genus and the differentia of the species. We might say that the species *human* belongs to the genus *animal* and is differentiated from other species of this genus by the possession of *rationality*. A human is therefore a rational animal – this is the set of defining features which comprise the essence of what it is to be a human.

It seems that Aristotle is of the opinion that this is the only truly scientific method of classification and that for any particular entity, the system can yield only *one* correct result. He appears to think that it is not plausible that there are two different but equally satisfactory ways of classifying a particular entity and that this is because being correct in matters of classification involves reflecting the natural order of things. So, for instance, in *De Partibus Animalium I*, Aristotle outlines the way in which the natural world ought to be divided, saying things like, "...it is necessary to divide by privation..",[10] "It is the differentia in the matter that is the species",[11] "...one should divide by what is in the being, and not by the essential accidents..."[12] and "...one should divide by opposites".[13] He also tells us that, "...we should if possible say that because this is what it is to be a man, therefore he has these things;

[9] *Categories*, 3^a 38-3^b 5.
[10] *De Partibus Animalium I*, 642^b 21.
[11] *De Partibus Animalium I*, 643^a 24.
[12] *De Partibus Animalium I*, 634^a 27-28.
[13] *De Partibus Animalium I*, 634^a 32.

for he cannot be without these parts."[14] and "They ought to say that the animal is *such*, and to speak about that – what it is and what kind of a thing..."[15] Further evidence is to be found in the *Metaphysics*, where Aristotle tells us that, "...the essence is what something is... Therefore there is an essence only of those things whose formula is a definition.",[16] "Nothing, then, which is not a species of a genus will have an *essence* – only species will have it...",[17] "Each thing and its essence are one and the same in no merely accidental way..."[18] and "...the essence is not produced; for this is that which is made to be in something else by art or by nature or by some capacity."[19][20]

Aristotle's account employs an objectivist metaphysics and an objectivist epistemology. He believes that the world in itself falls into a number of classes and that membership in these classes is governed by possession of certain essential or defining characteristics. Furthermore, he seems to believe that this division of the world is unique. As for epistemology, for our definitions and classifications to be correct, Aristotle appears to be of the opinion that they must reflect the classes and defining features inherent in the world. To have knowledge of a thing or to give the meaning of a term is to give the appropriate class together with the defining characteristics attached to the class. Knowledge and meaning therefore comprise reflections of metaphysics and no credit is given to the human beings who have knowledge and grasp meaning, nor to their cognitive processes.

John Locke, in Book Three of *An Essay Concerning Human Understanding*, drew a distinction between the real and nominal essence of things, the real essence being, "...the being of anything, whereby it is what it is. And thus the real internal, but generally (in substances) unknown constitution of things, whereon their discoverable properties depend..." and the nominal essence being, "...that abstract idea which the general, or sortal...name stands for."[21] Locke was of the opinion that this inner real essence was the cause of the phenomenal properties of material things. However, we can never be acquainted with the real essences of material things, since they are not visible to the naked eye and so we are forced to group things together on the basis of

[14] *De Partibus Animalium I*, 640^{a} 33-35.
[15] *De Partibus Animalium I*, 641^{a} 16-17.
[16] *Metaphysics*, 1030^{a} 3-7.
[17] *Metaphysics*, 1030^{a} 11-12.
[18] *Metaphysics*, 1031^{b} 17-18.
[19] *Metaphysics*, 1033^{b} 6-7.
[20] This characterisation of Aristotle is based on Staniland's (1972) account.
[21] Quotations from Locke, 1690/1964, p.270.

their phenomenal similarities. Locke also doubts, if we *were* able to discover real essences, that they would correspond to the nominal essences and kinds which we have set up. He says, for instance, "When we come to examine the stones we tread on, or the iron we daily handle, we presently find we know not their make, and can give no reason of the different qualities we find in them. It is evident the internal constitution, whereon their properties depend, is unknown to us. Therefore we in vain pretend to range things into sorts, and dispose them into certain classes under names, by their real essences, that are so far from our discovery or comprehension."[22]

Thus we can see that Locke was an advocate of objectivist metaphysics but not of objectivist epistemology. He was of the opinion that the entities of the world fall into unique natural classes on the basis of possession of real essences. This is quite independent of how human beings cognise and attribute structure to the world. Since he believed that these essences were invisible to the naked eye, he did not adhere to an objectivist epistemology. However, had he lived to see the invention of the microscope and the advances of science, we can imagine that he might have begun to argue for an objectivist epistemology.

Finally, Wittgenstein, in his *Tractatus*, was interested in language as a means of representing how things are in the world. He tells us that the world is the totality of facts – the existence of certain states-of-affairs. Facts are more and less complex, but in the last instance, there are so-called atomic facts which cannot be further broken down into simpler facts and which are mutually independent. Thus, "Even if the world is infinitely complex, so that every fact consists of infinitely many states of affairs and every state of affairs is composed of infinitely many objects, there would still have to be objects and states of affairs."[23] Our language is construed as corresponding isomorphically to the facts of the world. Thus, propositions are more and less complex, but in the last instance there are so-called atomic propositions which correspond to atomic facts. For instance, "The simplest kind of proposition, an elementary proposition, asserts the existence of a state of affairs."[24] Wittgenstein employs the metaphor of a picture, with language as the picture and reality as the thing pictured. The picturing metaphor depends upon the notions of a one-to-one correspondence between the elements of the picture and the thing being pictured and

[22] Locke, 1690/1964, p.287.
[23] Wittgenstein, 1961, paragraph 4.2211.
[24] Wittgenstein, 1961, paragraph 4.21.

on a common structure which is shared by language and reality. Thus, "We picture facts to ourselves." and, "A picture presents a situation in logical space, the existence and non-existence of states of affairs."[25]

In the *Tractatus*, we can see that Wittgenstein employs an objectivist metaphysics. He is of the opinion that the world has a detailed structure in and of itself and there is some indication that this structure is unique, since it comprises atomic facts which are basic and cannot further be analysed (presumably these at least are unique) and from which more complex facts are constructed. He says, for instance, "If all true elementary propositions are given, the result is a complete description of the world. The world is completely described by giving all elementary propositions, and adding which of them are true and which false."[26] He also employs an objectivist epistemology – by using the picturing metaphor, it is clear that Wittgenstein believes that in order to represent how things are in the world, language must mirror how the world is and must do so isomorphically and so exactly. The logical structure of language is identical with that of the world. The reason why language works is that it is anchored in reality by means of logical isomorphism. No attempt is made to explain how atomic propositions 'hook onto' the world nor how human beings grasp the meaning of the propositions which they use.

Contemporary Philosophy of Language.[27]

In his paper, "On Sense and Meaning", Gottlob Frege makes the distinction between 'sense' and 'reference'. He reasons that signs, such as words or linguistic expressions, have certain senses by means of which they refer to or pick out certain objects in the world – their references. For example, we might say that 'bachelor' means 'unmarried man' (sense) and this can pick out a particular individual or individuals (reference). According to Frege, a sign's sense is something universal and public – accessible to all users of the language. He says, "The sense of a proper name is grasped by everybody who is sufficiently familiar with the language or totality of designations to which it belongs..."[28] The sense of a sign is to be distinguished sharply from the idea which an individual person might have of that sign. We are told that, "...the idea which we have in that case is wholly subjective; in between lies the sense, which is indeed no longer subjective like the idea, but is yet not the object

[25] Wittgenstein, 1961, paragraphs 2.1 and 2.11.
[26] Wittgenstein, 1961, paragraph 4.26.
[27] This section is greatly indebted to Johnson's (1987) account of objectivist philosophy.
[28] Frege, 1952, pp.57-8.

itself."[29]

In order to ensure that sense can be public and universal in the way that is required, Frege distinguishes three quite separate realms – the physical; the mental, comprising subjective and personal mental representations; and the realm of thought, comprising all objective senses, numbers, propositions, concepts and functions. He believes that by distinguishing this last realm of thought from the perceived subjectivity and variability of individual and personal mental representations (the mental), the objectivity and uniformity of sense is ensured. He claims, for instance, that:

> "Such an idea is often imbued with feeling; the clarity of its separate parts varies and oscillates. The same sense is not always connected, even in the same man, with the same idea. The idea is subjective: one man's idea is not that of another. There result, as a matter of course, a variety of differences in the ideas associated with the same sense. A painter, a horseman, and a zoologist will probably connect different ideas with the name 'Bucephalus'. This constitutes an essential distinction between the idea and sign's sense, which may be the common property of many people, and so is not a part or a mode of the individual mind. For one can hardly deny that mankind has a common store of thoughts which is transmitted from one generation to another."[30]

Frege's account is clearly objectivist with regard to epistemology. On his view, the relationship between signs and objects in the world is an entirely objective one; he believes that there is some sort of direct linkage between (public) sense and objects or states-of-affairs in the world. This linkage is not considered to be mediated by anything like human cognition or understanding. In fact, human mental representations are dismissed as entirely subjective, hence irrelevant for matters relating to objective and universal sense. It is not made clear how human beings grasp a sign's sense nor is it made clear in what way senses connect to states-of-affairs in the world.

There are a number of contemporary semantic theories which are based upon the Fregean framework. One such example is model-theoretic semantics. Philosophers who are associated with this project are people such as David Lewis and Richard

[29] Frege, 1952, p.60.
[30] Frege, 1952, p.59.

Montague.[31] Model-theoretic semantics aims to give a precise account of how abstract symbols gain meaning via their correspondence to states-of-affairs in the world. The central notion here is that of a (mathematical) model which takes the form of a set-theoretical structure which comprises entities and sets of those entities. It is assumed that the actual and all possible worlds can be placed in one-to-one correspondence with a model so described. In order for this to be achieved, the further assumption is made that reality comprises entities with fixed properties and relations holding between them at any given time. Properties are placed in one-to-one correspondence with the set of entities possessing those properties and relations are placed in one-to-one correspondence with the set of entities standing in those relations. So, the idea is that there is a correct model of the actual world which simply comprises entities and sets of entities. Given such a model of the world, it is argued that abstract symbols (seen as meaningless in themselves) acquire meaning via their correspondence to entities and sets in this model.

It can therefore be seen that set-theoretical semantics is objectivist in both its metaphysics and its epistemology. As regards metaphysics, we have the assumption that the world in itself exhibits an inherent structure, quite independent of how we as human beings cognise, come to understand or interact with that world. As regards epistemology, model-theoretic semantics gives an entirely objectivist account of meaning. Meaning is again seen as something quite independent of those who grasp and use that meaning, since it is accounted for in terms of a direct unmediated relationship between abstract symbols and elements of models of the actual and possible worlds. Also, correct reasoning is equated with manipulating symbols in accordance with the logical (set-theoretical) structure of the model, which is seen as a reflection of the logical structure of the world – knowledge therefore involves a passive reflection of that structure. David Lewis states, for instance:

> "I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and second, the description of the psychological and sociological facts whereby a particular one of these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two topics."[32]

[31] See, for instance, Dowty, Wall and Peters, 1981 and Lewis, 1972.
[32] Lewis, 1972, p.170.

Another example of a semantic theory based on Frege's proposals is Davidsonian semantics. Davidson argues that to give an adequate theory of meaning is simply to give a theory of truth – he explains meaning by reference to truth. So, in order to give a definition of truth, one must give necessary and sufficient conditions for every sentence in the language and to specify these truth conditions (or to specify the state-of-affairs that would have to hold in the world in order for one such sentence to be true) is to give the meaning of that sentence. He talks about:

> "...the obvious connection between a definition of truth of the kind Tarski has shown how to construct, and the concept of meaning. It is this: the definition works by giving necessary and sufficient conditions for the truth of every sentence, and to give truth conditions is a way of giving the meaning of the sentence. To know the semantic concept of truth for a language is to know what it is for a sentence – any sentence – to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language."

In order to give truth conditions for sentences in the language, Davidson argues that it is necessary to employ a recursive theory which illustrates how larger truth-preserving units can be constructed out of smaller units which have truth conditions attached. He says, "...a theory of meaning for a language *L* shows "how the meanings of sentences depend upon the meanings of words" if it contains a (recursive) definition of truth-in-L."[33]

Again, it can be seen that Davidson's programme is broadly objectivist. He treats meaning as some kind of direct and unmediated link between sentences and states of affairs in the world. He makes no attempt to explain how smaller meaning-units link onto the world, nor to explain how humans grasp this meaning. It is assumed that all such human concerns are irrelevant to giving a correct theory of semantics, which he believes should be concerned only with truth conditions or with reflection of the world's inherent structure.

Putnam and Kripke.

I want to devote a special section here to the Objectivism espoused by Putnam and Kripke, since their work (mostly that of Putnam) is discussed at length in later chapters of this thesis. Their work deals specifically with questions of natural kinds

[33] Quotations from Davidson, 1967, p.310

and essentialism, but they also take on board many of the more general commitments of objectivist metaphysics and epistemology.[34]

Putnam's aim in his paper, "The Meaning of "Meaning"" is to challenge two components of earlier theories of meaning. These are (1) the idea that knowing the meaning of a term is simply a matter of being in a certain psychological state and (2) the notion that a term's meaning determines its extension. In short, he is aiming to show that the meaning of (in particular natural kind terms) depends much more upon the true or real nature of things than earlier theories allow.

Two things will only count as members of the same kind for Putnam if they possess the same nature or same "important physical properties". To determine whether they do, we must first determine what language users are referring to when they use a particular term and we must then uncover the nature of the referent of that term. For instance, we can determine that when speakers of English talk about 'water', they are referring to a clear, tasteless, odourless liquid which is to be found in pools, lakes and waterfalls. We (or specifically, our scientists) can then discover that the nature of that substance or its important physical property is being H_2O. This means that all liquids which are referred to as 'water' must possess the property of being H_2O and if not, it will be incorrect to term them water. This will be so even if these liquids are superficially similar to water. Superficial similarity is not enough to make them water. The only thing that is adequate to make them water is that they are H_2O, since that is the nature of water, that is what water is. As Putnam puts it, "Once we have discovered that water (in the actual world) is H_2O, *nothing counts as a possible world in which water isn't* H_2O."[35] For Putnam, then, the nature or essence of a natural kind is a cross-world relation which is cashed out in terms of important physical properties which are discovered (or discoverable) by science:

> "...we can understand the relation *same*$_L$ (same liquid as) as a cross-world relation by understanding it so that a liquid in world W_1 which has the same important physical properties (in W_1) that a liquid in W_2 possesses (in W_2) bears *same*$_L$ to the latter liquid."[36]

Putnam demonstrates that it is important physical or microstructural properties which

[34] It should be noted that the opinions attributed to Putnam here come from his early strong realist phase. He has since distanced himself from this position, arguing instead for theories such as his internal realism, which is discussed in the next section.

[35] Putnam, 1975, p.233.

[36] Putnam, 1975, p.232.

comprise the essence of natural kinds by means of the famous Twin Earth thought experiment. We are invited to imagine a place called Twin Earth which is exactly similar to Earth. In particular, Twin Earth has rivers, waterfalls and lakes which are full of a liquid which the Twin Earthians call 'water' and which has all the surface characteristics possessed by Earthian water. However, on closer inspection, it is found that Twin Earthian water does not have the microstructure H_2O, but a different microstructure – XYZ. Putnam therefore concludes that the Twin Earthian liquid is not water (it is not of the same kind as what we call 'water'), since it does not exhibit the requisite essence or nature.

Finally, Putnam makes the point that questions of natural kinds and their essences are quite independent of what human beings think or believe. Their existence is real and human beings simply discover this reality. So, for instance, Putnam asks us to imagine the Twin Earthian situation back in 1750, before the microstructure of water was known. He argues that *even though we would have had no way of telling Earthian and Twin Earthian water apart*, still, "...the extension of the term 'water' was just as much H_2O on Earth in 1750 as 1950; and the extension of the term 'water' was just as much XYZ on Twin Earth in 1750 as in 1950."[37] Likewise, Putnam argues that although our methods of identifying gold are now much more advanced than they were in ancient Greece, still the extension of 'χρυσὸς' in ancient Greek is the same as the extension of 'gold' in modern English. Therefore, even though the Greeks might have called various pieces of metal (which did not have atomic number 79) 'χρυσὸς', they would have been mistaken in doing so. These pieces of metal do not fall within the extension of 'gold' in modern English and neither do they in ancient Greek, reasons Putnam. This is because what comprises the essence of a natural kind is quite independent of human knowledge, beliefs and interests – it is rather a matter of the way reality is.

Although the origins of Kripke's arguments concerning natural kinds, as stated in his paper, "Naming and Necessity", are somewhat different from those of Putnam,[38] the general conclusions that he comes to concerning natural kinds are broadly similar. In particular, he aims to show that natural kind terms acquire meaning in virtue of their reference, in virtue of the real nature of those entities for which they stand.

[37] Putnam, 1975, p.224.

[38] He wants to emphasise, for instance, the similarity which he sees between proper names and natural kind terms.

Kripke is of the opinion that, in general, natural kind terms have their reference fixed by some kind of initial baptism. So, 'gold', for instance, might be defined as the substance which is instantiated by these items here.[39] Items which have similar superficial characteristics to those in this original set will then also be labelled 'gold'. With the advent of science, however, things change slightly. Once the reference of the kind has been fixed, Kripke believes that it is the job of science to determine what the real essence of the set is and so to determine which are the true and which are the deviant members of the set. As he puts it, "...science attempts, by investigating basic structural traits, to find the nature, and thus the essence (in the philosophical sense) of the kind."[40] Once the essence of the kind has been discovered in this way, it is impossible, according to Kripke, that the kind not have that essence (that gold not be the element with atomic number 79, for example) because *having that essence is just what it is to be a member of that kind.* To have atomic number 79 is simply to be gold; that is the *nature* of gold.

Kripke illustrates the fundamental nature of essence with a number of examples and thought experiments. The point of all these is to show that reference fixing tends always to be a matter of superficial properties – those entities which are superficially similar will initially be grouped together in the same kind. However, it is not these superficial properties which determine what it is to be a member of that kind – this is always a matter of deeper hidden properties which are amenable to scientific discovery. Hence, all sorts of situations can be constructed where the superficial characteristics of the kind turn out to be quite different from those which were originally supposed, yet the kind still remains in tact, since what defines the kind is its hidden essence and and not its surface properties. So, Kripke suggests that we could discover that gold is in fact not yellow, but blue – we have all been deceived by some sort of optical illusion. He argues that in this case it would not be announced that there is in fact no gold, but that although we supposed gold to be yellow, we were mistaken – it is in fact blue. Kripke also argues that if we came across entities with the same surface characteristics but quite different internal characteristics from members of a particular natural kind, these entities would not belong to that kind because they would not possess the requisite nature or essence. He therefore argues that we might discover some animals which look just like tigers, but which in fact have the internal structure of reptiles. These animals would not be tigers because they

[39] He allows for an 'almost all' qualification, which covers the situation where most of the items are gold, but a few are only superficially similar to gold and are not, in fact, gold.

[40] Kripke, 1972, p.330.

are not of the same species as tigers – they do not possess the same internal characteristics.

Finally, Kripke, like Putnam, makes it clear that natural kinds and essences are not influenced by what we, as human beings, know or think. We can therefore correctly speak of the essence of a kind without knowing what that essence is. As Kripke puts it, "I think this is true of the concept of tiger *before* the internal structure of tigers has been investigated. Even though we don't *know* that internal structure of tigers, we suppose...that tigers form a certain species or natural kind. We then can imagine that there should be a creature which, though having all the external appearance of tigers, differs from them internally enough that we should say that it is not the same kind of thing. We can imagine it without knowing anything about this internal structure – what this internal structure is. We can say in advance that we use the term 'tiger' to designate a species, and that anything not of this species, even though it looks like a tiger, is not in fact a tiger."[41]

It is clear that the Putnamian and Kripkean accounts are objectivist in nature. In particular, they embrace the objectivist notions of essentialism and natural kinds. Certain essential properties define the nature of natural kinds and both the properties and kinds are objective or real aspects of the natural world. Furthermore, it is assumed that these kinds and their essences are fixed by metaphysics, by the way the world is – both Kripke and Putnam point out that once the essence of a kind has been discovered, it is simply not possible that the essence could be anything other than what has been discovered. In other words, both accept the notion of reality being uniquely structured in a particular way. This reality is fixed and independent of any human input.

Putnam and Kripke also clearly adhere to an objectivist epistemology. To give a correct account of natural kinds and their nature is to mirror the structure which is inherent in the world. Their work clearly shows that they wish to acknowledge the role that the real world plays in the characterisation of these terms. Putnam is keen to overturn the notion that intension or meaning determines extension and instead to demonstrate that a difference in extension automatically results in a difference in meaning. He thereby implies that meaning is purely a matter of what is the case in the world and is unaffected by human interests, actions or understanding. No effort is made to explain how human beings grasp that meaning. Likewise with Kripke; a

[41] Kripke, 1972, p.318.

statement telling us what the essence of a kind is, as discovered scientifically, will be a necessary truth – the essence of a kind is simply what it is to be of that kind, it supplies us with the meaning of the appropriate kind term, but this meaning is something which is specified by the world or metaphysics alone. There is a direct and unmediated link between the term and the world and it is by referring directly to things in the world that terms acquire their meaning. Meaning is construed as something quite independent of human beings, their interests and their understanding.

Problems with Objectivism.

As we have seen, there are two components of Objectivism; objectivist metaphysics and objectivist epistemology (cognition and language). There are accordingly two angles from which we can argue against Objectivism. We might argue that its metaphysics is wrong and we might argue that its epistemology is wrong.

If one were to argue against objectivist metaphysics, one would have to argue against the notion of an independently structured world consisting of entities, properties, relations and categories. In particular, I wish to argue against the classical view of categorisation, especially in relation to natural kinds. This is a task which I undertake in detail in Chapters Six and Seven of this thesis by means of a mixture of case studies (Chapter Six) and philosophical thought experiments and arguments (Chapter Seven). Briefly, my point is that the way in which scientists categorise objects in the natural world is not purely a matter of the way the world is. Rather, scientists must make a *choice* as to which properties are to count as the salient ones for matters of categorisation. Which properties are chosen will determine what counts as a relevant similarity for grouping entities together for the purposes of categorisation. The notion lying behind this is one of a world which is conceptually extremely rich and which can therefore support more than one classification of its entities. Different scientists will want to highlight and downplay different aspects of reality and they are able to do this by taking different aspects or properties of real-world entities as salient for matters of classification. This means that the properties which are judged to be essential for any one class will depend, to a certain extent, on the aims and purposes of the scientist doing the classification (and so may vary, dependent upon those aims and purposes) and, of course, what is taken as essential dictates where the boundaries of the class fall. In other words, members of so-called natural kinds together with the

essences and boundaries of these classes are not purely a matter of metaphysics, but involve some degree of human input. If this picture is correct, then objectivist metaphysics must be mistaken.

However, in this section, I want to discuss what is wrong with objectivist epistemology – with the objectivist account of language, cognition, meaning, rationality and knowledge. Although I initially separate this from metaphysics, the two are closely interwoven and by looking at what is wrong with the epistemology, we provide ourselves with an entry into and possible solution of the problems with the metaphysics.

The Putnam / Johnson Strategy.

Recall that the objectivist believes that language and thought are simply a matter of mirroring the logical fixed structure of the world. Language acquires meaning by referring to entities and states of affairs in the world. In order to have knowledge, we must reflect the structure of the world. Rationality becomes transcendent and impersonal. All this implies that there exists a God's Eye view of the world which is attainable by, but quite independent of, rational human beings. This comprises an entirely neutral view of reality, a kind of 'view from nowhere'.

Hilary Putnam takes issue with this notion of a God's Eye view of reality, arguing that such a view is impossible to attain and so meaningless. This approach is adopted and extended by Mark Johnson (sometimes in collaboration with George Lakoff).[42] I will now discuss and develop the Putnam / Johnson strategy.

Putnam's Internal Realism.

Putnam's general aim in his *Reason, Truth and History* is to go beyond the philosophical positions of Subjectivism and Objectivism and to formulate a new position which lies between these two, taking the best elements from each. As he puts it, "In the present work, the aim which I have in mind is to break the strangle hold which a number of dichotomies appear to have on the thinking of both philosophers and laymen. Chief among these is the dichotomy between objective and subjective views of truth and reason."[43]

Putnam characterises the position which he calls "metaphysical realism" as, "...the

[42] Johnson, 1987; Lakoff and Johnson, 1980. See also Lakoff, 1987.
[43] Putnam, 1981, p.ix.

world consists of some fixed totality of mind-independent objects. There is exactly one true and complete description of 'the way the world is'. Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things."[44] He refers to this position as the "externalist perspective", since it implies the existence of a God's Eye point of view, a way in which we can stand outside the universe and offer the uniquely correct description of it. This view is clearly parallel to the one which we have been calling Objectivism.

In opposition to metaphysical realism, Putnam launches an alternative view, which he labels the "internalist perspective" (or internal realism). The core of this view is that "*...what objects does the world consist of?* is a question that it only makes sense to ask *within* a theory or description."[45] In other words, talk of a God's Eye perspective is meaningless. No such perspective can exist. To talk of the way the world is in itself, independent of the way it is cognised or described makes no sense. Given a particular theory about or description of the world as a basic starting point, it is possible to talk about the way the world is, but without some kind of prior perspective, it is just impossible to talk.

Putnam tells us that truth on this view is, "...some sort of (idealized) rational acceptability – some sort of ideal coherence of our beliefs with each other and with our experiences *as those experiences are themselves represented in our belief system...*"[46] The important point to extract here is that Putnam entirely rejects the notion of correspondence between our experience and the world as it is in itself, independent of our cognition and language. This is because it simply makes no sense to talk of having access to a world in itself. If this is the case, then a comparison between our experience and an independent world is impossible. Our experiences of the world are, by definition, mediated. Our experiences automatically include something of us, as human beings. This something is likely to be a mixture of our human physiology, together with our beliefs, training and cultural inheritance which have been inculcated in us since birth and which shape the way we cognise, think and speak about the world. Even at the level of experiences, then, it makes no sense to speak of having a God's Eye view of the world – our view must, by definition, be a human, not a godly one.

[44] Putnam, 1981, p.49.
[45] Putnam, 1981, p.49.
[46] Putnam, 1981, p.50.

A further crucial point made by Putnam is that the theory of internal realism does not boil down to an out-and-out relativism in which anything goes. As the name implies, internal realism still has its feet firmly *in reality*. It is important to realise that under no circumstances is the existence of an external world denied. Rather, we constantly live in and interact with the world, we constantly have cognitive inputs from it. However, as explained above, those inputs logically cannot be direct and unmediated, they cannot be from God's point of view. Thus our experience, knowledge, theory and description of the world is continually constrained by that world, but it is constrained by that world as *we* experience it:

> "...it does deny that there are any inputs *which are not themselves to some extent shaped by our concepts*, by the vocabulary we use to report and describe them, or any inputs *which admit of only one description, independent of all conceptual choices*. Even our description of our own sensations...is heavily affected...by a host of conceptual choices."[47]

What follows from this is that talk of having access to the world as it is in itself is meaningless. If it is accepted that our experiential inputs cannot be unmediated, that we cannot speak of experience independent of any kind of conceptualisation, then the notion that there is any God's Eye perspective to be attained makes no sense. This is why any theory of truth which involves correspondence of our concepts or linguistic items to a mind-independent reality is a non-starter for Putnam. In order to fix such a correspondence, both sides of the correspondence must exist – something which the internal realist denies. One of the consequences of this, as Putnam points out, is that some internal realists believe that, since there is no such thing as one description of reality independent of all cognition and language, there is more than one possible (true) theory or description of the world. He says that there are, "...various points of view of actual persons reflecting various interests and purposes that their descriptions and theories subserve."[48] Lakoff gives a good illustration of this point – take a chair; there are various ways in which that chair can be correctly described. From the molecular perspective, it is seen as a collection of molecules and so not as a "single undifferentiated bounded entity". From the perspective of wave equations in physics, there is no chair, only wave forms. From a lay human's perspective, it is simply a chair – a single bounded entity. In every case it is granted that something exists in the external world, but this thing can be described in very different ways from different

[47] Putnam, 1981, p.54.
[48] Putnam, 1981, p.50.

perspectives or from within different conceptual schemes. Interestingly, it is only from the lay human perspective that the chair is considered as a single object – what constitutes an object and how the world should be divided into objects varies between theories of description. Each perspective is correct, but only from *within* the relevant conceptual scheme or theory of description, hence there can be more than one true description of the world. Furthermore, it would be incorrect to attribute absolute priority to any one of these differing but true descriptions.[49]

For the internalist, then, meaning cannot be a matter of a word or concept correctly referring to some entity in an objectivist world (since we do not have experience of the objectivist world). Rather, meaning is a matter of a particular group or community of language users using signs in a particular way to refer to entities *as those language users experience those entities*. Putnam is of the opinion that, "'Objects' do not exist independently of conceptual schemes. We cut up the world into objects when we introduce one or another scheme of description. Since the objects and the signs are alike *internal* to the scheme of description, it is possible to say what is what."[50]

Likewise, with kinds or classes of objects. The objectivist argues that the world in itself falls into entities which themselves fall into classes. To say that something is of the same kind as something else is therefore to say that really, objectively or metaphysically, the two things belong together in the same class. Yet the internal realist has shown that it is meaningless to talk of direct access to the world in this way. Putnam therefore points out that, "...'of the same kind' makes no sense apart from a categorial system which says what properties do and what properties do not count as similarities."[51] In other words, if we cut the world up into objects, we must also cut it up into kinds. By isolating, defining and labelling an object, we *ipso facto* designate the kind to which that object belongs. All other objects which are similar to that object in certain relevant ways will belong to the same kind. But this is not simply a matter of reflecting the inherent structure of an objectivist world, rather it is a matter of interpreting and shaping up our irreducibly human experiences. It is a matter of our deciding in what respects two things must be similar in order to belong together in the same class and this decision will be made relative to a conceptual scheme or theory of description which prefers certain properties over and above

[49] Lakoff, 1987, p.262.
[50] Putnam, 1981, p.52.
[51] Putnam, 1981, p.53.

others for the purposes of classification.

The above is a fairly brief sketch of Putnam's internal realism which he offers as an alternative to metaphysical realism (or Objectivism).[52] I will now turn to Mark Johnson's development of this position, known as experientialism or experiential realism.

Johnson's Experientialism.

> "We are concerned here with how *real human beings reason* and not with some ideal standard of rationality. We are concerned with what *real human beings grasp as meaningful.*"[53]

Johnson's account can be seen as a development of Putnam's internal realism with special reference to its implications for a theory of meaning. He is particularly keen to stress the human nature of meaning. He argues that linguistic items do not acquire meaning by latching onto and mirroring the inherent logical structure of the world, as Objectivism would have it. Rather, they acquire meaning by being used by human beings in certain ways for certain purposes.

He makes the further important point that Objectivism gives no account of what it is for human beings to grasp or to understand the meaning of terms. Objectivists tend to treat human understanding as though it were uninteresting and straightforward – they claim that it just *is* the grasping of the relation between linguistic items and the structure of the world (or between sentences and truth conditions or their conditions of satisfaction). On this account, the notion of grasping remains unanalysed. Of course the reason for this is that Objectivism provides, by definition, a dehumanised account of meaning. Its primary aim is to posit an ultimately logical, rational account

[52] It is perhaps worth noting that in more recent years the emphasis of Putnam's views has shifted slightly, although his general aim to find a mid-point between metaphysical realism and relativism remains. In particular, he makes much more of the point that the world is as it is, irrespective of the interests of people describing it, although he continues to assert that there are a variety of descriptions which can be true of some portion of that world, that there is no such thing as a description which reflects no interests at all and that giving a description is never a matter of simple copying. He also rejects the notion that the world is a product of the human mind, expressing regret that, in *Reason, Truth and History*, he suggested that that the world was, to a certain extent, mind-dependent. (See Putnam, 1994, p.448.) Putnam's modified views emphasise points which I am extremely anxious to stress in my own analysis of categorisation, such as that differing categorisations must be based upon patterns or similarities which really do cut across the natural world and that we are not entitled to divide up the world in any way that we like. To this extent, then, Putnam's modified views are perhaps even more in tune with my own than was his original exposition of internal realism.

[53] Johnson, 1987, p.11.

of how language can relate to the world, unsullied by such messy factors as human psychology, vagueness and ambiguity. It considers such human factors to be irrelevant to an account of meaning, since meaning ought to be objective and human factors simply introduce subjectivity.

Johnson, however, argues against this that, "...between symbols and the world, there falls understanding."[54] In other words, he feels that it is impossible to relate symbols to the world without some sort of account of how human beings interact with and conceptualise their world. This is almost exactly the same point as Putnam makes when he claims that asking what the world or its structure is like only makes sense from within a particular conceptual scheme or theory of description. In order to relate symbols to the world, you must have a particular scheme or description as a basic starting point, you must have an internal perspective, since an external God's Eye view of the world and how symbols relate to it is a meaningless concept. So, for Johnson, understanding requires being situated firmly *within* the world. He claims that this places emphasis on the "dynamic, interactive character"[55] of understanding and meaning. Meaning is no longer something fixed or written in stone as the objectivist would have it, rather meaning becomes an active and interactive process. Meaning emerges as a result of the continued interaction of humans with the environment in which they are always situated. Likewise, grasping a meaning is an active process. It is not simply a case of reflecting the inherent structure of the world correctly by means of symbols, rather it involves being situated in that world and understanding how and why your fellow humans collectively use symbols to structure and makes sense of their environment. Understanding for Johnson is, "...a historically and culturally embedded, humanly embodied, imaginatively structure event."[56] It is only through an appreciation of this cultural, biological and social process that we are able to arrive at a full and proper recognition of how symbols relate to the world or of how meaning arises.

Thus, for Johnson, a theory of meaning requires a theory of understanding. One of the major reasons for this, according to Johnson, is that words cannot have meaning in and of themselves. What makes a word meaningful is that it is used by a community of language speakers to refer to something in a particular way. Words are only meaningful for speakers who *use* those words. Although we might say that a

[54] Johnson, 1987, p.175.
[55] Johnson, 1987, p.175.
[56] Johnson, 1987, p.175.

word means such-and-such, what we are actually indicating by this is that the word is *used by a community of speakers to mean* such-and-such. This is really a further extension of the point that words do not acquire meaning simply by reflecting states of affairs in an objectivist world – rather, meaning is something which is mediated. It is human use and so understanding which mediates between words and the world. This use and understanding arises from the situation of human agents within their environment and not from their adopting a God's Eye or external perspective. Without this kind of mediation, words are literally meaningless. Objectivists simply assume that sentences map onto the objectivist world by virtue of the fact that there exist states-of-affairs in the world which satisfy those sentences (conditions of satisfaction). However, Johnson makes the point that this totally ignores *why* those sentences have the conditions of satisfaction which they do. His answer is that conditions of satisfaction only count as such because they are conditions of satisfaction *for us* and this is so relative to our experience and understanding (our internal perspective) of the world in which we are situated and with which we interact. Again, using Putnam's terminology, it only makes sense to say that sentences have particular conditions of satisfaction from within a theory of description or conceptual scheme.

Further to this, Johnson argues that, "The meaning of a symbol stems from the imposition upon it of a certain intentionality, which is *always a matter of human understanding.*"[57] Intentionality refers to the capacity that representations such as concepts, words, sentences and images have to be about something outside themselves. He makes the point that this is first and foremost a property of our mental states and only secondarily a property of sentences, concepts etc. by dint of the fact that *we impose* intentionality upon them. So, his argument is that these kinds of symbols are intentional – they are about other things, yet intentionality is always dependent upon the human mind. A mental image, word, concept will be intentional because it is *used by humans* to represent some element of their experience. Hence, for such symbols to be meaningful, it is a prerequisite that there exist creatures for whom they are meaningful. The way in which they acquire that meaning is by being used by those creatures to represent or relate to something beyond the symbol. This something will not be some element of the objectivist's universe, since, as Putnam has shown, access to such a universe is impossible. Rather it will be the experience and understanding which a community of symbol users has of the world in which

[57] Johnson, 1987, p.177.

they are situated and with which they are constantly interacting. As Johnson concludes, "...[meaning] is never merely an objective relation between symbolic representations and the world, just because there can be no such relation without human understanding to establish and mediate it."[58]

Again, it should be borne very strongly in mind that Johnson's account, like Putnam's, neither involves a rejection of realism nor allows for a type of anything goes relativism. Johnson is in no way denying that there exists an external world which is independent of us. This is in fact borne out by our continuous interaction with and participation in that outside world. He says, ""Things" outside us talk back to us, and proclaim their presence, with a very loud voice most of the time. And it is our being situated in relation to them, that assures us that we are realists."[59] He points out that it is our continuous interaction with our environment which has prompted us to come up with structures of understanding which enable us to function more-or-less successfully within that environment. If we are to function successfully, it is very clear that an anything goes relativism is simply not an option. If we fly an aeroplane into a mountain, we will wreck the craft and very likely be killed. This is the way in which the world continuously talks back to us. Any conceptual scheme which makes it O.K. to fly into mountains will prove to be a dead (!) loss pretty quickly. The point, however, that Johnson wants to make is that although the environment will of course place limitations on our conceptualisation of it, it is simply not the case that there is one God's Eye view of the world. We can still conceptualise that world in different ways, dependent upon our aims and purposes in interacting with it. A God's Eye view is meaningless, a myth and this is because it would require us to stand outside the environment, to adopt an external perspective. Yet, the nature of our humanity is that we are an integral part of our environment. What constitutes being human is operating within a world, it affecting you and you affecting it. By virtue of the very fact that we are human, our perspective is an internal one.[60]

George Lakoff, during discussion of these issues, makes another very important and insightful point.[61] He maintains that recognising that a God's Eye point of view is

[58] Johnson, 1987, p.178.
[59] Johnson, 1987, p.204.
[60] It is worth emphasising that even if some being could stand outside the world, there would still be no unique God's Eye description available to that being, simply because there are many potential ways in which the world can be (correctly) described. It is therefore only from *within* a particular conceptual scheme that one particular description can justifiably be singled out for special attention.
[61] Lakoff, 1987, pp.301-2.

impossible does not in any way involve denying that objectivity (rather than Objectivism) is impossible. He claims that being objective involves putting aside one's own point of view and being able to look at a particular situation from as many other points of view as possible. This involves being aware of one's own beliefs and the conceptual framework in which they reside, being aware of other points of view together with their conceptual frameworks and being able to look at and understand a situation from these other points of view. Lakoff goes on to emphasise that according to these standards, Objectivism actually involves a rejection of objectivity, since the objectivist claims that there is one and only one correct way of describing the world. This is the description which mirrors the inherent structure of the world. Any alternative system is dismissed out of hand.

Conceptions of Truth.

One of Putnam's aims in *Reason, Truth and History*, is to advance a new theory of truth in opposition to the traditional correspondence (or copy) theory of truth which tends to go hand in hand with versions of metaphysical realism. I believe, however, that this conception of truth, as he presents it, is confused and unnecessarily complex. At times, it even appears to smack of the metaphysical realism which he is aiming to oppose. I believe that given internal or experiential realism, we can stick to the notion of a correspondence theory of truth, so long as this is construed as correspondence within a particular context. This kind of approach is taken by Mark Johnson.

It is helpful to begin with a brief summary of what Putnam says about truth. He begins by claiming that an internal realist conception of truth is, "...some sort of (idealised) rational acceptability – some sort of ideal coherence of our beliefs with each other and with our experiences *as those experiences are themselves represented in our belief system* – and not correspondence with mind-independent or discourse-independent 'states of affairs'."[62] This, of course, makes sense, since, according to internal realism, the notion of a God's Eye context-free point of view is meaningless. If our true statements accord with anything, then it must be with our experience of the world and not with a God's Eye notion of the world. As Putnam puts it, "You can't single out a correspondence between two things by just squeezing *one* of them hard (or doing anything else to just one of them); you cannot single out a correspondence between our concepts and the supposed noumenal objects without

[62] Putnam, 1981, pp.49-50.

access to the noumenal objects."[63]

Putnam goes on to define "rational acceptability". He explains that what makes a statement or a theory rationally acceptable is coherence of theoretical beliefs with each other and with experiential beliefs and coherence of experiential with theoretical beliefs. He claims that our notions of acceptability and coherence are dependent on our humanity – our psychology, biology and inherited culture. On this picture, it is therefore possible to end up with two theories or conceptual schemes which are equally coherent and rationally acceptable and so to end up with more than one 'true' description of reality. The idea is that incompatible theories or descriptions may correspond equally well to or make equally good sense of *our* experiences of the world. A consequence of this, according to Putnam, is that the notion of scientific truth is dependent upon certain values. This is because, "*...truth is not the bottom line*: truth itself gets its life from our criteria of rational acceptability, and these are what we must look at if we wish to discover the values which are really implicit in science."[64] So, to claim that science is aiming to construct a true picture of the world is really a meaningless statement until you spell out the criteria of rational acceptability to which science adheres.

However, Putnam insists that to reject the metaphysical realist's conception of truth does not amount to *identifying* truth with rational acceptability – the two are not equivalent notions. This is because what is rationally acceptable may change over time and because what is rationally acceptable can be seen to be a matter of degree (whereas we see truth as an absolute notion – a statement is either true or false, not true to a certain degree). He therefore defines truth as an *idealisation* of rational acceptability. He says, "We speak as if there were such things as epistemically ideal conditions, and we call a statement 'true' if it would be justified under such conditions."[65] He acknowledges that we can never really attain epistemically ideal conditions or even be sure that we come anywhere close to them. He likens them to frictionless planes, explaining that these cannot be attained either but that they are of use to us because we can approximate them very closely. The implication seems to be, then, that we can approximate ideal epistemic conditions very closely too.

It is with this notion of idealised rational acceptability that, I feel, Putnam's theory

[63] Putnam, 1981, p.73.
[64] Putnam, 1981, p.131.
[65] Putnam, 1981, p.55.

runs into confusion. In the preface to *Realism with a Human Face*, Putnam attempts to explicate this notion further. He says that to claim that a statement is true in its context or conceptual scheme is to claim that, "*...it could be justified were epistemic conditions good enough.*"[66] or that it could be justified were epistemic conditions ideal. He then gives an example. An ideal epistemic condition in relation to Putnam's claim that there is a chair in his study would be if he was in his study with the lights on or with daylight coming through the window, with unimpaired eyesight, with a clear mind and if he looked to see if there was a chair there. He adds that it is possible to drop the word "ideal" altogether and simply say that there are better and worse epistemic situations with regard to particular statements. The above example would then represent a very good epistemic situation.

I think the basic intuitions behind Putnam's account of truth are correct. That is, I agree that a notion of truth which requires correspondence with a God's Eye notion of the world is untenable, since such a God's Eye view is impossible to attain and so meaningless. Putnam is right when he claims that what is required is some sort of agreement between our theorising about and our experience of the world. However, I think he gets himself into trouble over the notion of idealised rational acceptability. The problem is that he employs the notions of idealised rational acceptability and ideal epistemic conditions in order to explain the concept of truth but then emphasises that neither notion can actually be attained by us – each is a fiction. This suggests that we can never attain truth, that what we do attain cannot be counted as truth and that what we do have is in some way inferior to the unattainable ideal of truth. In fact, by holding on to this kind of notion of truth, Putnam appears to be reinstating a situation of the very kind which he was aiming to reject in writing *Reason, Truth and History*. That is, on the one hand, there is what is rationally acceptable for humans to believe which they utilise in their negotiation of the world and on the other hand there is truth (idealised rational acceptability) – some kind of abstract notion which is unavailable for human use. Idealised rational acceptability appears to play the same kind of role as the the God's Eye view of the world played in metaphysical realism. That is, it appears to be a construct which is entirely independent of any sort of human cognition or judgement – a kind of view from nowhere. Yet this is quite at odds with the kind of view which Putnam is trying to present. If my analysis of Putnam's idealised rational acceptability is correct, then by the lights of his own account, it should be rejected along with and for the same

[66] Putnam, 1990, p.vii.

reasons as the God's Eye view of the world.

Putnam's later claims in *Realism with a Human Face* in fact contradict his earlier position in *Reason, Truth and History*. To claim on the one hand that there are ideal but unattainable epistemic conditions and to claim on the other that there are better and worse epistemic conditions with regard to particular statements is not at all to claim that same thing. No one would deny the latter statement and I think people would agree that a statement made in good epistemic conditions is more rationally acceptable (and so more likely to be true) than one made in poorer epistemic conditions. However, good or very good epistemic situations are ones which *are attainable by humans* whilst, by definition, unattainable or fictitious epistemic conditions are not. Furthermore, the chair in the study example given by Putnam involves epistemic conditions which *are* possible to attain. It looks, then, as if there is some inconsistency between the two accounts and that Putnam's later version involves somewhat different claims from his earlier version.

I therefore think that there are better and simpler ways of presenting a conception of truth which accords with the basic intuitions of the internal realist account. The notion of truth presented by Mark Johnson fulfils this role.

Johnson, of course, employs much the same basic intuitions as Putnam; intuitions that a God's Eye view of the world is meaningless and that between symbols and the world there must fall human understanding. However, in relation to the question of truth, Johnson retains a realist intuition which Putnam abandons – the intuition that truth involves statements which correspond to states-of-affairs in the world.

Johnson claims that, "Truth-as-correspondence is still a workable notion only if it is not understood in the objectivist fashion, as requiring a God's-Eye-View of an external relation between words and the world."[67] He argues that given a particular conceptual scheme which divides the world into entities and kinds of entity, certain statements will correspond more accurately to that world than others. However, he qualifies, "...this "correspondence" will always be relative to our *understanding* of our world (or present situation) and of the words we use to describe it."[68] In other words, this correspondence is not a case of statements in themselves simply corresponding to states-of-affairs in themselves in one and only one way, since as

[67] Johnson, 1987, p.210.
[68] Johnson, 1987, p.203.

Johnson has earlier argued, this kind of unmediated mapping of statements onto the world is impossible. Rather, it is a case of statements which are employed by human beings for particular purposes corresponding to those same human beings' experience and so understanding of the world in which they are situated. Since correspondence is not absolute, it is feasible that different statements correspond in different ways, dependent on the aims and purposes of the humans who make those statements and have the experience to which they correspond.

The point is made that an absolute God's Eye notion of truth is not something which we require:

> "It doesn't really matter that we can't see the world through God's Eyes; for we can see the world through shared, public eyes that are given to us by our embodiment, our history, our culture, our language, our institutions, etc...it does mean that we can know that we are partially in touch with reality, not in the "one correct way" but in one or more of the possible ways in which Nature can be described. Thus, we can still preserve a notion of truth-as-correspondence, as long as it is contextually situated."[69]

Perhaps the most important part of this quotation is the last sentence. It simply makes no sense to define truth in an absolute way, given that a God's Eye notion of the world is impossible to attain. However, given a particular context, conceptual scheme or point of view on the world, it makes sense to differentiate between true and false statements. Later on in this thesis, it will become apparent that this conception of truth fits with scientific classification, since I illustrate that it is possible, on the basis of differing scientific theories, which embody different aims and purposes, to come up with competing definitions of what it is to belong to a particular kind and with competing classifications of particular entities, which highlight different real properties of the entities which are being classified. These competing definitions and classifications are equally scientific and fit (our experience of) the facts equally well. We therefore have a situation where competing statements are both true, given the differing theories or contexts in which they are embedded. As Johnson puts it,

> ""Accurately describing reality" is not a single, homogeneous purpose on a par with a purpose like making one's bed. "Describing accurately

[69] Johnson, 1987, p.211.

> how things are" is a shorthand for "finding descriptions of reality that work more or less well given our purposes in framing descriptions of reality."...Truth is always truth relative to a basic description and relative to standards of adequacy determined by our human purposes and the nature of our interactions with our environment."[70]

We can therefore see that Johnson appears to make a somewhat better job of dealing with truth than does Putnam, whilst preserving the same internal or experiential realist intuitions. He holds onto the important realist notion that truth should involve some kind of correspondence with states-of-affairs in the world, but manages to combine this with the notions that the God's Eye view of the world is meaningless and that statements cannot simply correspond to reality in a unmediated fashion.

Conclusions.

In this chapter, I have given a full account of the doctrine of Objectivism, discussing both the metaphysics and the epistemology invoked by such an account. We have seen that what is missing from this account is any mention of how human beings think about or conceptualise their world. According to Objectivism, reality comes with a preferred theory of description and this is how reality objectively is, from a God's Eye perspective. Knowledge, concepts, language and meaning simply reflect the inherent logical structure of the world. Any mention of how humans think about the world is seen as introducing a subjective element into an exclusively objective picture and is therefore unacceptable.

We have seen how Putnam challenges Objectivism by claiming that a God's Eye perspective of the world is unattainable. This is because all knowledge of the world requires some kind of prior structuring and conceptualisation by the organisms which possess that knowledge. He instead introduces a new doctrine called internal realism which claims that it is possible for humans to have knowledge of an independent external world, but emphasises that this knowledge is from an internal perspective. We are a part of this world and so the knowledge we have of it is from the inside, as a result of our interaction with it and is mediated by own experience and conceptualisation of it, which in turn is mediated by our biology, culture and history. Johnson extends Putnam's doctrine, arguing that meaning is not constituted by symbols simply latching onto bits of the objectivist world, but is an essentially

[70] Johnson, 1987, p.211.

human phenomenon which is constituted by a group of language users using symbols to refer to parts of their environment as they experience that environment. The Putnam/Johnson account reasserts the important role played by the human agent in matters of meaning, language and knowledge.

Chapter 3
Internal Realism: Criticisms and Analysis.

Introduction.

In this chapter, I examine some criticisms which have been levelled against the parts of Putnam's internal realism which I discussed in the previous chapter. Although these criticisms are aimed directly at Putnam, they can also be taken as criticisms of my characterisation of Objectivism and of Johnson's experiential realism, since these are extensions of Putnam's doctrine and so take on board the basic tenets of his position.

Secondly, I discuss how my account of categorisation tallies with and is dependent on conceptions of metaphysics and epistemology as described by internal realism. I also discuss how internal realism represents a half-way house between Objectivism and Subjectivism and how the empirical evidence which I make use of in my thesis is consonant with the basic tenets of this doctrine.

Criticisms of the Internal Realist Approach.

Hartry Field, in his paper, "Realism and Relativism", raises a number of objections to Putnam's position, as I have presented it.

Objection 1.

Field correctly characterises Putnam's account of metaphysical realism in terms of three statements. These are:

- Metaphysical realism$_1$ = "the world consists of some fixed totality of mind-independent objects".
- Metaphysical realism$_2$ = "there is exactly one true and complete description of "the way the world is"".
- Metaphysical realism$_3$ = "truth involves some sort of correspondence

relation between words or thought signs and external things and sets of things".[71]

Field then goes on to argue that he does not consider metaphysical realism$_2$ to be a consequence of metaphysical realism$_1$, that he believes metaphysical realism$_2$ to be false and that it, "should not be taken as a component of any sane version of realism".[72] He argues that it would be possible for there to be a true description of the world in a language completely alien to our own (it may use predicates with extensions which are not easily definable or not definable at all in our own language or it may even use referential mechanisms which are quite different from predicates). How are we to say that two such true, yet different, descriptions of the world could be the same?

I think that in response to Field, it is useful to look at some of the objectivist philosophers whom I discussed in the previous chapter. What we will see from this is that proponents of Objectivism or of metaphysical realism *do* tend to take on board the notion that there is "exactly one true and complete description" of the world. Perhaps this goes to show that, in general, versions of metaphysical realism are *in*sane!

Consider Aristotle and Wittgenstein. As we have seen, Aristotle is of the opinion that the world is divided inherently into species and genera – this is the natural order of things. Thus to give a correct classification of all the entities in the natural word is to reflect that natural order. For Aristotle, then, we can infer that this is the one true and complete description of the world. For Wittgenstein, the world comprises facts which can ultimately be broken down into atomic facts. The role of language is to represent the way things are in the world and therefore language comprises propositions which correspond isomorphically to facts and which share the same structure as those facts. For Wittgenstein, then, the one true and complete description of the way the world is involves a series of propositions which correspond exactly to the totality of facts which comprise that world. As he says, "If all true elementary propositions are given, the result is a complete description of the world. The world is completely described by giving all the elementary propositions, and adding which of them are true and which false."[73]

[71] Field, 1982, pp.553-554.
[72] Field, 1982, p.554.
[73] Wittgenstein, 1961, paragraph 4.26.

Or take Frege. As has already been discussed, Frege is determined to ensure the universal and objective character of sense and in order to secure this, he introduces the notion of a realm of thought. One of the supposed consequences of Frege's doctrine is that there is a direct correspondence between universal sense and reference (or states-of-affairs in the world). I suggest that it is natural to construe this realm of thought as equivalent to the one true and complete description of the world in virtue of the fact that for each and every state-of-affairs in the world, there should be a corresponding sense which will have objective and universal currency as a member of the realm of thought. The totality of senses will therefore count as the complete description of the world.

Likewise, with model theoretic semantics. The notion here is that *in virtue of the fact* that the world consists of a fixed totality of mind independent entities and relations between those entities, this world can be placed in one-to-one correspondence with a set-theoretical model. Again, it would seem natural to infer that the set-theoretical model will comprise the one true and complete description of the world and it is in virtue of corresponding to elements in that model that abstract symbols come to have a constant and universal meaning.

Or we can look at the Putnam/Kripke approach to natural kinds. Kripke, for instance, constantly reminds us that a member of a particular natural kind cannot fail to have the requisite essence, because having that essence is simply what it is to be a member of that kind. Thus, "And so, it seems to me, this would not be a case in which possibly gold might not have been an element, nor can there be such a case...Given that gold *is* this element, any other substance, even though it looks like gold and is found in the very places where we in fact find gold, would not be gold. It would be some other substance which was a counterfeit for gold."[74] The consequence of this is that it is *wrong* to place an entity in a natural kind if that entity lacks the requisite essence. Hence Putnam claims that it would be wrong to call something 'water' which lacked the chemical structure H_2O or to call something 'gold' which lacked atomic number 79, even if the 'deviant' entities were superficially similar to the 'real' members of the kind and even if we were not acquainted with the requisite essences and so had no means of telling the 'real' and 'deviant' entities apart. The reason why this would be wrong is that Putnam and Kripke believe that the world in itself is divided into kinds on the basis of essential or defining properties, quite

[74] Kripke, 1972, p.320.

independent of any human activity. They therefore believe that the one true and complete description of the kinds and essences of the world is the one which reflects these kinds and essences, as they exist in nature. Any description which fails to do this will be false and incomplete.

So, I maintain that most of the metaphysical realists whom I have discussed *do* believe that there is a unique, complete and true description of the way the world is. Whether this position is sane or insane, true or false, is quite another question, but nevertheless, it is a position which has (frequently) been held and therefore Putnam is acting quite legitimately in arguing against it.

Given that it is the case that a number of metaphysical realists do believe that there is one complete and true description of the world, we can now look at Putnam's own response to Field's criticism. He argues that one of the ways in which we might make sense of the claim that there is one true and complete description of the world, is to assume that metaphysical realism$_1$ is true – that there is a fixed totality of mind-independent objects. This means that there will be a definite set of individuals *I* of which the world consists and a definite set of all properties and relations which pertain to those individuals, *P*. We can now imagine an ideal language which has a name for each entity in *I* and a predicate for every member of *P*. As a result of isomorphism with the world, this language and its set of true sentences can be construed as a unique, complete and true description of that world. Putnam therefore concludes that, "the natural way of understanding "metaphysical realism$_2$" involves assuming "metaphysical realism$_1$"".[75] In other words, the two doctrines are mutually supporting.

Another way of putting Putnam's point might be this. If you have a metaphysical realist who accepts that the world consists of a fixed totality of mind independent objects and who accepts that truth involves some sort of correspondence between symbols and external things, then it would seem obvious to assume that this metaphysical realist accepts the notion that there is one complete and true description of reality. To give a complete and true description of the world is to give a list of *all* the statements which are true of the world, in virtue of corresponding to *each and every* element of or state-of-affairs in that world.

As we have seen, a number of philosophers appear to adhere to both metaphysical

[75] Putnam, 1990, p.31.

realism$_1$ and metaphysical realism$_2$, treating the one as a consequence of the other. It is clear, then, that this sort of position has, in the past, been offered as a component of realism. Whether such philosophers are misguided in proposing this kind of position is a different matter. However, it would seem that when Field argues that concepts very different from our own might be used to describe the world, he is in fact buying into one of the tenets of *internal* realism – that the world can admit of more than one true or correct description – rather than producing a criticism of metaphysical realism, as characterised by Putnam. He therefore needs to provide stronger arguments concerning (a) why he disagrees with metaphysical realism$_2$ and why metaphysical realism$_2$ is not entailed by metaphysical realism$_1$ and (b) why he is not therefore prepared to accept Putnam's internal realism.

Objection 2.

Field takes issue with Putnam when he claims that, "*what objects does the world consist of?* is a question that it only makes sense to ask *within* a theory or description."[76] Field correctly assumes that by this, Putnam means that, "different correct theories or descriptions will answer the question differently."[77] He then goes on to look at Putnam's suggestion that Maxwell's electromagnetic field theory and action-at-a-distance formulations of electrodynamics are both equally good and coherent theories which fit our experiential beliefs equally well, despite the fact that they are metaphysically incompatible since the one claims that 'fields' exists and the other does not. As Field puts it, the two theories differ in ontology.

Field argues that this example neither suggests that objects are mind- or theory-dependent, nor does it pose a problem for the correspondence theory of truth. He claims that this is because advocates of field theory need not deny that the behaviour of particles can be explained without fields, just so long as you, "introduce enough complexity into the basic equations of motion."[78] Likewise, advocates of action-at-distance theories need not deny the existence of fields, but, "merely refrain from asserting their existence."[79] He continues, "Most of us recognise that more exists than one need assert the existence of: it is rarely to the point to assert the existence of undetached rabbit parts as well as of rabbits...there is clearly no way to draw anti-realist consequences merely from the fact that two equally good theories could differ

[76] Putnam, 1981, p.49.
[77] Field, 1982, p.555.
[78] Field, 1982, p.555.
[79] Field, 1982, p.555.

in their existence claims."[80]

I think there are a number of things which need to be said here. Firstly, we should note that Putnam claims that these two theories are *metaphysically* incompatible. What this means is that the idea that both theories can be correct is a problem for the *metaphysical* realist, as Putnam has characterised that position. To recap, the metaphysical realist, as characterised by Putnam, believes that (a) there is a fixed mind-independent totality of objects, that (b) there is exactly one true and complete description or theory of the world and that (c) truth involves a correspondence between words or thoughts and external things and sets of things. Clearly, then, this example *is* problematic for Putnam's metaphysical realist. By dint of (a), either fields must exist or they must not, yet the point of this example is that this cannot be answered conclusively – whether fields exist or not will depend on what theory you are working within. By dint of (b), there cannot be more than one theory which correctly describes the same phenomenon, yet in this example, we have two theories which explain the same phenomenon equally well.

So, Field's claims that more exists than need be asserted will be of small comfort to Putnam's metaphysical realist, since this is just what he believes to be impossible. If the world consists of a fixed totality of objects and if there is just one true theory of the world, then all that needs to be asserted (by the true theory) will be all that *can* exist. Conflicting theories and incompatible objects are just not acceptable to the metaphysical realist. Of course, Field has already stated that he disagrees with Putnam's characterisation, but (a) I have already illustrated that his disagreement is unfounded on the basis of what objectivist philosophers have actually claimed and (b) *given* Putnam's characterisation, the field/particle example *is* problematic for the metaphysical realist.

Furthermore, this example also poses a problem for the correspondence theory of truth within Putnam's characterisation of metaphysical realism. Given that the world consists of a fixed totality of individuals plus properties and relations pertaining to those individuals, what makes a word or thought or statement true is that it corresponds one-to-one with individuals/properties/relations in that world. If the statement, "physical events = particles acting at a distance" is true, then it must be a fact in the external world that there exist particles which act at a distance but if there are actually fields in the external world which mediate the action of these particles,

[80] Field, 1982, pp.555-556.

then this statement will be false. Both states-of-affairs cannot be actual, since the objects of the world are a fixed set, there is only one true description of that world and truth involves unique correspondence to the fixed objects of the world. Field's attempt to fix the problem in terms of what can be asserted simply doesn't work on Putnam's characterisation of metaphysical realism. The idea that we, as human beings, might have some choice over what (not) to assert and that we could be correct on both counts if we chose to assert two conflicting things simply does not fit the metaphysical realist's picture of the world as uniquely determined, quite independent of any human activity.

However, these two incompatible theories do not pose a problem on the internal realist picture. In fact, Field's notion that more exists than we need assert actually fits this picture very well. As Putnam himself points out, "...[it] is not that correspondences between words or concepts and other entities don't exist, but that *too many* correspondences exist. To pick out just *one* correspondence between words or mental signs and mind-independent things we would have already to have referential access to mind independent things."[81] And, "If truth is not (unique) correspondence then the possibility of a certain pluralism is opened up."[82] The point is that for Putnam, the notion of us having access to a unique God's Eye point of view simply makes no sense. We are human, therefore our access to the world will always be mediated by our humanity – by our culture, our biology, our aims, interests and so on. This opens up the possibility that we can make sense of our experiences of the world in different, yet equally coherent ways, as in the field/action-at-a-distance example. The fact that one theory asserts the existence of fields and the other does not is not problematic, because the internal realist has not taken on board the notion of the world as structured completely independent of us nor has he taken on board the notion of a God's Eye point of view, together with its attendant notions of a unique description of the world and so of truth as being unique correspondence to that world. In contrast, for the internal realist, it is important that within a theory, our theoretical beliefs cohere with one another and with our experiential beliefs. Since the field and action-at-a-distance theories fit this requirement, they are equally good and coherent theories. Putnam's internal realism therefore very much allows for the suggestion that more exists than we need assert – it allows for the fact that fields might exist, yet whether we assert that they exist or not will depend upon which

[81] Putnam, 1981, pp.72-73.
[82] Putnam, 1981, p.73.

theory we are working within.

Finally, it is worth saying something very briefly about Field's assertion that anti-realist consequences cannot be drawn from the field/action-at-a-distance example. I do not think that Putnam ever had any intention of drawing anti-realist consequences from this example. It should be borne in mind that the position he advocates is called internal *realism* and he makes it very clear throughout that we, as human beings, are a part of reality and so have continuous experiential inputs from that reality. What he *does* deny, however, is that it makes any sense to say that these experiences are from a God's Eye point of view. Rather, they are our *human* experiences which will, by definition, be affected and so shaped by all sorts of human factors. What Putnam *does* aim to do, then, is draw anti-*metaphysical* realist consequences from the field/action-at-a-distance example. He aims to attack the God's Eye view of the world, of descriptions of the world and of truth, since he sees these as comprising an untenable form of realism. To conflate this with the aim to draw anti-realist consequences, however, would be mistaken and would be to overlook entirely the great insights of Putnam's position.

Objection 3.

Thirdly, Field examines Putnam's statements that, "objects do not exist independently of our conceptual scheme"[83] and that, "objects themselves are as much made as discovered"[84]. He interprets Putnam as meaning something like this – we might have had conceptual schemes which were so different that we did not think it terms of dinosaurs, but in terms of quite different entities. Field offers undetached dinosaur parts as a suggestion here. However, he complains that this is simply not enough to make Putnam's metaphor work, reiterating the point made in the criticism above that more exists than we need to assert the existence of – hence both the dinosaur and the undetached dinosaur parts exist, whether anyone asserts that they exist or not, whether anyone thinks in terms of them or not. Furthermore, Field is concerned that as the book goes on, "one becomes more and more tempted to think that Putnam does in some literal sense believe that we created the dinosaurs. Again and again...he tells us that facts depend on human cognitive values..."[85]

I am not sure that Field has fully understood Putnam here. According to Putnam, yes

[83] Putnam, 1981, p.556.
[84] Putnam, 1981, p.557.
[85] Field, 1982, p.557.

we do 'make' objects, but not literally. I think the best way to understand Putnam's point is in terms of kinds. Objects do not exist independently of conceptual schemes, because it is only under a particular conceptual scheme that two things can be identified as being of the same kind (of object). To pick out and label an object is to associate particular properties with that object in virtue of which other objects of the same kind can be identified. However, the world does not come ready-packaged with those properties picked out as the salient ones, rather it is our conceptual schemes which pick out objects and the way in which they achieve this is to mark out certain properties as salient in virtue of which we can say objects are similar and so of the same kind. As Putnam himself puts it, "...'of the same kind' makes no sense apart from a categorial system which says what properties do and what properties do not count as similarities. In *some* ways, after all, anything is 'of the same kind' as anything else."[86] And, "If, as I maintain, 'objects' themselves are as much made as discovered...then of course objects intrinsically belong under certain labels; because those labels are the tools we used to construct a version of the world with such objects in the first place."[87]

The easiest way to illustrate this is with an example. Coincidentally, one of the case studies which I discuss in Chapter Six examines how we originally acquired the term 'dinosaur'. Very briefly, this case study explains how the category 'dinosaur' was created by one man, Richard Owen. Prior to Owen, what we now call dinosaurs had been viewed as enormous reptiles, called 'fossil lizards'. This was based on fossil jaws and teeth which indicated similarities to extant lizards. However, Owen gave these creatures their own taxonomic rank – that of *Dinosauria* – based on anatomical peculiarities of the sacrum, ribs and extremities and on their enormous size, which served to distinguish them from living lizards as well as from Mesozoic marine lizards. What we have here is an example of an object being 'made' as much as discovered. Whether something counts as a dinosaur or not will depend on whether you are working within a conceptual scheme which counts dinosaurs as a separate class of things. If it does, then there will be properties associated with that class in virtue of which objects can be identified as instances of the class. If not, then the objects which we call dinosaurs will fall in with some other class on the basis of a different set of salient properties.

So, of course we don't literally make dinosaurs. Dinosaurs – those particular bits of

[86] Putnam, 1981, p.53.
[87] Putnam, 1981, p.54.

matter – existed at some point in the past no matter what we think or say. However, the fact *that* something existed does not necessarily wholly determine exactly *what* existed. From my synopsis of the case study, it can be seen that humans had a hand in making these entities *dinosaurs*, as opposed to giant lizards. That is, although these entities existed at some point in the past and exhibited certain properties, no matter what, which of these properties we, as human beings, find or take to be salient is a different matter. And which properties we do take as salient will determine whether these entities are dinosaurs or giant lizards, as has been illustrated above. When Putnam speaks of facts depending on human cognitive values, then, he is not arguing that we literally created the dinosaurs, but that how we divide up and categorise our environment depends, at least partially, upon properties and factors we take to be salient, on what is of interest *for us*. Thus Putnam says:

> "Take the sentence 'the cat is on the mat'. If someone actually makes this judgement in a particular context, then he employs conceptual resources – the notions 'cat', 'on', and 'mat' – which are provided by a particular culture, and whose presence and ubiquity reveal something about the interests and values of that culture, and of almost every culture. We have the category 'cat' because we regard the division of the world into *animals* and *non-animals* as significant, and we are further interested in what *species* a given animal belongs to. It is *relevant* that there is a *cat* on that mat and not just a *thing*. We have the category 'mat' because we regard the division of inanimate things into *artefacts* and *non-artefacts* as significant, and we are further interested in the *purpose* and *nature* a particular artefact has. It is relevant that it is a *mat* that the cat is on and not just a *something*. We have the category 'on' because we are interested in *spatial relations*."[88]

As I argued in connection with the former criticism, Field's point that both dinosaurs and undetached dinosaur parts exist whether anyone thinks in terms of them or not is not something that Putnam would deny. In fact, this is just the type of situation which internal realism allows for but which metaphysical realism, as characterised by Putnam, cannot allow for, for the reasons discussed above.

Finally, Field's observation that the notion that our conceptual schemes might have been quite different is not nearly enough to make Putnam's metaphor (that we make

[88] Putnam, 1981, pp.201-202.

objects as much as discover them) work is partially justified, I think. That is, the verb 'to make' tends to carry connotations of physically making and so was probably a bad choice of term on Putnam's part. Furthermore, I think that Putnam subsequently realised this – as noted in footnote 52, in more recent work, he expresses regret at having talked about mind dependence in relation to the world in *Reason, Truth and History*. However, I think it is unfair to Putnam to take sentences such as this out of context and them criticise them in isolation. Given the totality of what Putnam says, it is clear that he does not mean that we literally create the world. Rather, that our cognition of the world involves all sorts of predispositions, interests and choices which reflect characteristics of us, as human beings. This, however, is a picture which the *metaphysical* realist is at pains to avoid.

Analysis and Implications.

In the previous chapter, I looked at the objectivist doctrine and its implications for both metaphysics and epistemology (cognition and language). I also looked at the major problems with this account and at the alternative accounts which have been put forward by Putnam and Johnson. What we saw primarily was that Objectivism makes no mention of any role played by human beings. They are treated as merely incidental to such an account and metaphysics, language, meaning, rationality and knowledge are defined as transcendent and are in no way considered to be either dependent upon or influenced by human action or cognition.

Putnam has shown that the God's Eye perspective which is required by Objectivism is just meaningless. To say anything at all about the world, one must have a basic conceptual scheme or theory of description as a starting point. To say anything about the world, some form of initial conceptualisation and division into objects and categories is required. To deny this and to claim that there is one uniquely correct description of the world from an external perspective makes no sense, it is an impossibility. Putnam and Johnson therefore dispense entirely with the notion of a God's Eye external perspective. They acknowledge that human beings do not stand outside reality, but are a part of it. By the very fact that we are human, we belong within our environment. Since we are a part of our environment, we are continuously interacting with it. This means that both environment and humanity mutually affect one another. Since we, by definition, have an internal perspective on the world, Putnam and Johnson point out that our experience of that world will always be

mediated and structured by our conceptual schemes and theories of description. It is acknowledged that different but equally correct descriptions of the world (which are motivated by different conceptual schemes) are possible. As Lakoff puts it, "[this kind of approach] focuses on the way that we make sense of reality by functioning within it. The internalist perspective acknowledges the contribution of our conceptual schemes to our understanding of our real experiences in a real world."[89] Most importantly, the internal perspective draws our attention back to the human perspective. It is emphasised that human beings play a very active role in coming to an understanding of the world of which they are a part. They are not simply passive mirrors of the world – they actually attribute structure to and so, up to a certain point, shape that world.

However, despite the fact that human beings are accorded an active role on the internalist account, basic notions of realism are preserved throughout. Both Putnam and Johnson are committed wholeheartedly to the existence of a world outside the human mind. We experience this world constantly and the world therefore constrains our conceptualisation of it. Although different conceptualisations of it might be possible, this does not mean that we can say whatever we like about the world. It constrains what we are able to say about it and to ignore these constraints is to risk being frustrated by it, to risk being unable to negotiate it.

Having considered these basic points, we are now in a position to touch on the implications of such an account for more specific areas which are relevant to this thesis.

Categorisation.

The main body of my thesis concentrates on the categorisation by human beings of 'natural' objects (ones which occur in nature, rather than being manufactured by humans) and one of the main aims of the thesis, as stated above, is to show that classification of these objects involves more than simply reflecting the way the world is (or metaphysics), that it also involves an element of choice or decision by the person doing the classification and this will be dependent on the aims, purposes and commitments of that person. I aim specifically to show that this applies to scientists whose job it is to classify the natural world. Objectivism tends to consider science as its ally and so it is very interesting to note that the evidence from science actually

[89] Lakoff, 1987, p.263.

sheds doubt upon, rather than supports, the objectivist thesis.

By looking at Putnam's internal realism we can come to see how, theoretically, such a non-objectivist account of scientific classification is likely to be correct. As he explains, we divide the world up into kinds relative to particular conceptual backgrounds or commitments and it is these backgrounds which determine which similarities will and which will not count as relevant as the basis upon which to classify natural entities. He points out that things can be similar to other things in an infinite number of ways and so in order to answer the question of which kinds exist in the world, we must be coming from and so be motivated by a certain perspective which determines what is and what is not to count as relevantly similar. In other words, when a scientist classifies, he is operating with an *internal* perspective. It is logically impossible for him to stand outside the world, to operate with an external perspective, since interaction with and participation in the world is what constitutes being human. It is impossible to stand outside the world whilst describing it because, by that very act of description, a particular perspective is being adopted.

It is important to realise that although I am claiming that humans have some hand in determining how natural objects are classified, I am not trying to argue that we are able to classify in any way that takes our fancy. To do so would very quickly prove fruitless for humanity. As with the earlier example of the plane and the mountain, it would be unhelpful (at the very least!) if we had no mechanism in place to differentiate dolphins from sharks, when the one is friendly and the other a deadly killer or if we had no mechanism by which to differentiate edible from poisonous mushrooms when the one can kill and the other nourish you. Of course the environment will place limits upon our classifications of it – as Johnson says, the world continually speaks back to us. The point I am arguing for is that the world does not come with one preferred theory of description. It is therefore up to us to build a description of the world which tallies with our aims and purposes. Different aims and purposes may result in differing descriptions, each of which will satisfy a different need. However, one of the overriding purposes of any description or classification must be that it enables us to negotiate, cope with, or interact successfully with our environment. If there is no classification system in place which allows us to distinguish the harmless from the potentially lethal elements of our world, the chances of our long-term survival within that world will be drastically reduced. On a slightly less fatalistic, but parallel, note, if gold (being the element whose atomic

number is 79) is of high value in our society, then it is important that we have a system of classification which makes the distinction between surface appearance and internal structure a salient one and which thereby enables us to distinguish 'true' gold from substances bearing a surface similarity to gold.

Another way of making the same point is to again stress that this kind of account of classification does not involve abandoning the basic doctrine of realism. Of course there is a world alive and well outside of ourselves and consequently that world affects and impinges upon us in both profitable and dangerous ways. To ignore this is to ensure our own rapid demise. We must respect the ways of the world when negotiating it but still there is room for us to leave our mark on it, as it leaves its mark on us. To deny this is to deny our own reality and the very real effects which we have upon the world which are displayed when we negotiate and cope with it. It is important to realise that we are an integral part of our environment and that consequently there is a continuous mutual process of interaction between it and us.

We can begin to see, then, how the Putnam/Johnson approach naturally invokes the kind of theory of categorisation which I want to put forward. This section is intended to illustrate how my views on categorisation fit in with a much wider network of issues. The views themselves will be developed in detail during later chapters of this thesis.

Objectivism and Subjectivism: a Half-Way House.

The next point which needs to be made is that both Putnam's and Johnson's account and my treatment of categorisation represent mid-points between the two extremes of Objectivism and Subjectivism. As I mentioned right at the beginning of the account of internal realism, Putnam sees himself as opting for a middle ground between this dichotomy and so as going beyond a fierce debate which has dogged philosophy for so long.

Subjectivism is a doctrine which is diametrically opposed to Objectivism. It denies basic realism. There is no way the world is – it lacks any inherent structure. Hence all structure is imposed solely by us – we effectively construct reality. Any conceptual scheme is as good as the next. There is no such thing as truth. It is not possible to compare alternative conceptual schemes – they are incommensurable. And so on. Contrary to Objectivism, with Subjectivism, the emphasis is exclusively on humans and their construction of the world.

Mark Johnson points up the fundamental mistake which is made by both objectivists and subjectivists. Both camps make the mistake of driving an enormous wedge between the world and human beings, treating each element as self-contained and entirely separate from the other. The problem then becomes one of welding the two together, of determining which influences which. The objectivist opts to put all his money on metaphysics or the way the world is. Human beings then simply become passive mirrors reflecting the world's inherent and logical structure. Language, concepts etc. are no more than reflections of that world. They are transcendent and in no way affected by the creatures who use or have them. The subjectivist, on the other hand, opts to put all his money on the human factor. He claims that the world is nothing apart from human cognition and imposition of structure. Without this, the world is undifferentiated, a mish-mash. Language and concepts are human artefacts by means of which humans express the reality which they have created.

The picture painted by internal realism moves entirely away from this split scenario. The basis of this theory is quite different. It starts with the presupposition that human beings are a *part* of the environment in which they operate. It is impossible to drive a wedge between these two and to view humans as independent, since one of the things which constitutes being human is being a creature which operates *within* a particular environment. Our nature is constituted as much by our surroundings as by things internal to us. Once this kind of viewpoint has been adopted, you are no longer left with the problem of mapping one thing onto another, of deciding how the one is related to the other. Since the organism is considered to be an integral part of the environment, you can acknowledge the fact that both organism and environment affect one another – causality and influence work both ways. Humans interact continuously with their environment, learning how to negotiate and cope with it and so leaving their mark upon it. Likewise, the environment talks back to us; there are certain things we can never be successful at doing, simply because that is the nature of the world vis-à-vis us. As Johnson puts it, "...our structured experience is an organism-environment interaction in which both poles are altered and transformed through an ongoing historical process."[90]

In my opinion, this is a vitally important insight. It really does seem to provide a means by which to move beyond the locked battle of Objectivism versus Subjectivism, whilst preserving the best elements from each theory. Our notions of

[90] Johnson, 1987, p.207.

realism are preserved, whilst humans are allowed some input into the way the world is. It also enables us to move beyond the somewhat Medieval characterisation of man existing over and above the natural world as some kind of rational copy of God in contrast and superior to other non-rational creatures and inanimates. It acknowledges that man is just as much a part of nature as anything else and is not in some way distinct from it. With my account of categorisation, I aim to put these insights into practice.

In later chapters of this thesis, I seek to challenge classical philosophical views of categorisation, which claim that there is an absolute and unblurred dividing line between metaphysics and epistemology. Such views make the claim that which category an entity really belongs to is purely a matter of metaphysics (or the way the world is) and that it has nothing whatsoever to do with epistemology (or the ways in which human beings construe the world as being). The claim is then made that all psychological data concerning categorisation is quite irrelevant as any indicator of category structure in the world, since such data is purely a matter of epistemology and so only tells us about the ways in which we access our concepts and nothing about the categories of the world. These views are clearly based on Objectivism, with all its attendant assumptions of an independent logically structured world.

I argue, in contrast to the traditional philosophical positions, that we can draw no such clear dividing line between metaphysics and epistemology in the realm of categorisation and that the distinction between the two is in fact very blurred. The reasoning behind this is that humans have some input into the categorisation of the the world, that they play a much more active role than merely mirroring the metaphysics. In other words, categorisation involves a *mixture of metaphysics and epistemology*.

This argument is, in fact, an alternative way of putting across the internal realist's contention that there exists a mid-point between Objectivism and Subjectivism which goes beyond and is more fruitful than either of these two polar opposites. By arguing that the line between metaphysics and epistemology is blurred and that categorisation involves some kind of human input, I am in fact denying that one can intelligibly separate human beings from the environment in which they function, as a part of that environment. To claim that categorisation is purely a matter of metaphysics and that the only 'correct' categorisations are the ones which reflect the unique and inherent structure of the world, is to claim that a God's Eye view of the universe is both

intelligible and possible. As internal realism has shown us, such a perspective is impossible to attain. We are a part of our environment and so such a context-free perspective is unintelligible. Since the notion of one uniquely correct description of the world has been ruled out, it is possible that the world can support differing and conflicting classifications of its entities.

Internal realism makes the point that both organism and environment continuously interact, each one affecting and constraining the other. This is parallel to my argument that the line between metaphysics and epistemology is blurred in the domain of categorisation. We are constrained in categorising the entities in the world by the fact that we need to be able to negotiate and make sense of the ways of that world. However, there is still more than one way in which we can successfully and productively categorise those entities, dependent upon our aims, purposes and interests. We affect the world and the world affects us in the process of an ongoing mutual interaction. It is for this reason that categorisation cannot simply be a matter of metaphysics, that it must also be a matter of epistemology and that the dividing line between the two is extremely fuzzy. In fact, in order to be able to talk about a metaphysics at all, it is a fundamental prerequisite that one have an epistemology which provides the context in which to talk about the metaphysics, which provides the background against which the question, 'Which categories exist in the world?' is intelligible and so answerable.

Another way of making the point is this. We, as human beings, are actually a part of metaphysics or the way the world is. Since we are a part of the world, our experience and conceptualisation of the world around us is part of the fabric of the way that world is. The different ways in which we conceptualise our environment represent the different ways in which the environment is capable of being. We truly are interactive and fully functioning components of the environment and this is why it is impossible to draw a clear distinction between metaphysics and epistemology.

Empirical Evidence.

At various points in this thesis, I make use of empirical evidence. This is so in Chapter Four where I give an account of the history of the psychology of categorisation, laying emphasis upon the demise of the classical view as a result of the overwhelming body of experimental evidence showing that the majority of human concepts simply do not exhibit a classical type structure. This is also the case

in Chapter Six where I look at case studies of how scientists classify entities in the natural world. These case studies strongly support the argument that even at the level of science (considered to be the most objective level of human knowledge), classification is not merely a matter of reflecting the inherent structure of the world, but also involves a proportion of human input, which is guided by the interests, aims and theoretical commitments of the scientists doing the classifying.

In discussing the question of empirical evidence, we need to bring to mind the key elements of Objectivism once again. It needs to be remembered that Objectivism covers both metaphysics and epistemology. That is, on the one hand, the world in itself (metaphysics) is held to have an internal logical structure. It divides naturally into entities, properties and relations and these entities fall into classes on the basis of possession of common properties. On the other hand, it is this internal metaphysical structure which determines our epistemology. In order to gain knowledge of the world, we must mirror or reflect this internal structure. Any talk of human beings affecting their environment or metaphysics, of their understanding of the world playing any part in their conceptualisation of it is entirely ruled out under this kind of account. The objectivist is therefore going to be uninterested in empirical evidence concerning our concepts or the means by which we classify entities. The reason for this is that what people *actually do* is of no relevance to his theory. All that is of relevance for him is right there in the world itself and what people do only serves to confuse and muddy the situation, their vagueness and ambiguity getting in the way of the clarity and logical nature of metaphysics. The only human factor which might be of interest to the objectivist is when humans 'get it right', when they 'correctly' mirror the structure of the world, but this is still determined by and purely parasitic upon the metaphysics, since it is but a reflection of that metaphysics. The objectivist is not, then, interested in the humanity but in the metaphysics which is projected through rational humanity. This kind of attitude can be witnessed in a couple of papers written by the philosopher Georges Rey,[91] in which he discusses (or rather dismisses) the wealth of empirical evidence amassed which illustrates the fact that most human concepts do not exhibit a classical type structure. He claims that by taking this kind of empirical evidence seriously as telling us anything about conceptual structure, we are conflating metaphysics and epistemology. He claims that this evidence might tell us a good deal about people's access to their beliefs, but it tells us nothing about what concepts are, what their identity conditions are or what

[91] Rey, 1983 and Rey, 1985.

conditions are necessary to make someone a competent user of a particular concept. He continues, "The metaphysical – to my mind the only serious – notion of concept has consequences not only for the usual philosophical and scientific questions of what *is* what, but also for questions about what people *believe* to be what. People, when they classify things, are as much engaged in metaphysics as the philosophers and scientists who are paid to do so..."[92] In other words, as far as our concepts and categorisations go, metaphysics is what is relevant and metaphysics alone determines how we conceptualise the world. This is so as much for non-experts as for experts. We, as human beings, make no important contribution whatsoever to the way our world looks or is. Anything (such as psychology) which looks at how humans think, reason and conceptualise is quite irrelevant for telling us about how either the world or how our concepts are structured.

However, in this chapter we have seen what the problems with an objectivist account of both metaphysics and epistemology are. Putnam has argued that the kind of God's Eye perspective required by the objectivist makes no sense. There are no such things as unmediated experiences of the world. All experience requires some prior conceptualisation (some basic conceptual scheme or theory of description). It is only against this kind of prior background that it makes sense to talk about what types of object and what kinds of classes exist in the world. Johnson has extended Putnam's remarks to cover meaning. Since unmediated experiences of the world are impossible (since the notion of a God's Eye view is empty), it cannot be the case that meaning is constituted by some sort of automatic latching of concepts or words onto the world, quite independent of humans. Johnson argues that meaning is constituted by a group of language users using words to mean or refer to entities or states-of-affairs in the world as those language users experience them. Meaning is therefore dependent on what humans do and think, it is not something independent or free-floating, as the objectivist would have it. We have seen, therefore, that Objectivism needs to be replaced by an internal perspective. It needs to be acknowledged that humans constantly interact with and form a part of their world. Their experience of that world is mediated by their culture, tradition, history and biology and is from the inside – it emerges as a result of their involvement *within* the environment.

On this new internal realist perspective, humans are given the respect they deserve. It is acknowledged that we affect the world as well as that the world affects us (the

[92] Rey, 1985, p.298.

interaction is mutual) and therefore that the way we divide up and classify entities in that world cannot be purely a matter of metaphysics – we must contribute in some way to our perception, experience and conceptualisation of our environment, since we form a part of that environment. For the internal realist, then, empirical evidence is bound to be of relevance, since he is genuinely interested in how humans think, reason and conceptualise and he is interested in this because he believes that, to a certain extent, humans shape their environment. Epistemology is, for the internal realist, relevant to metaphysics. There is no clear demarcation between the two, since each one has an affect upon the other. It is clear from the psychological evidence that the majority of human concepts do not exhibit a classical type structure. The internal realist will accept that this is a statement about the structure of our concepts and so a statement about the way in which we impose structure upon our world since, for him, concepts are about the way in which *humans conceptualise their own environment*, they are not simply a reflection of the internal logical structure of the world. This evidence tells us *how people think* and this is directly relevant to the way in which they shape their world. Likewise, case studies of scientific classification throw doubt on the objectivist picture. They show us that even science (traditionally considered to support Objectivism) is affected by human aims, theories and commitments and that scientific classification is not purely a matter of metaphysics, but involves human choice and judgement. This is what the internal realist would expect and these kinds of study give very strong support to his theory.

Conclusions.

In this chapter, I have examined a number of criticisms levied against Putnam's internal realism. We have seen, however, that these criticisms tend to rest on misunderstandings of the basic tenets of this doctrine and can, in general, be shown to be problems for the metaphysical, rather than for the internal, realist. Putnam's doctrine, therefore, remains unscathed.

It is in the context of internal realism that my account of categorisation comes to life. I claim that categorisation is as much a matter of epistemology as of metaphysics and that the human agent plays a much more active role than simply reflecting the way the world is. He must make decisions and choices and these will be determined by his prior aims, interests, theories and commitments. It is possible to come up with differing, but equally valid categorisations of entities and the world is able to support

these differing categorisations. This is not to say that we can categorise in any way we like. The world constrains our division of its entities, but it does not come with only one unique theory of description. In order to appreciate such a theory of categorisation, one must first of all understand the problems with and reject Objectivism, whilst replacing it with a more human-oriented account of meaning, language and concepts. Putnam and Johnson between them achieve this and it is this kind of internal realist doctrine which serves as a philosophical backdrop to my account of categorisation.

I have tried to emphasise that my account of categorisation takes the middle road between Objectivism and Subjectivism. Categorisation is not purely a matter of the way the world is, nor is it purely a matter of human whim and fancy. Rather, it is an admixture of the two; the world imposes limitations on our categorisation of it, yet how we choose to categorise that world will partly depend upon factors of human interest. This middle of the road approach is also characteristic of internal realism. Both Putnam and Johnson want to move beyond Objectivism and Subjectivism, combining elements of the two. They succeed in preserving basic realism, whilst according importance and respect to human thought and action.

Finally, I have attempted to explain why empirical evidence is relevant and important for my account of categorisation. We have seen that the psychological evidence suggests that our concepts do not have a classical (objectivist) structure and that scientific case studies indicate that classification in science is not only a matter of metaphysics, but also involves active human input. What is clear is that human thought, concepts and practice do not seem to fit the objectivist account of epistemology. This is, in fact, fundamentally important, but it is only likely to be judged as such if one is an internal realist. Objectivists will simply dismiss such empirical evidence, arguing that what human beings think and do is irrelevant. The world is as it is whatever humans do and if they were supremely rational, if they could see through God's eyes, then their concepts and thought processes would conform to the objectivist account of epistemology. The internal realist, however, dismisses the God's Eye perspective as an impossibility. On his account, concepts and language are human tools for thinking about and communicating human (not God's) experiences of the world. He will therefore interpret such evidence as really telling us something about conceptual structure and meaning from the only possible perspective – the internal one.

An internal realist might further argue that empirical evidence concerning human concepts and categorisation actually tells us something about metaphysics as well as epistemology. This is not because epistemology is a mirror image of metaphysics, but because it only makes sense to talk about the way the world is from within the context of a particular theory of description or conceptual scheme, from within a particular epistemology. Therefore, how we mentally represent and structure our concepts, how or by what means we categorise entities in the world will be intimately tied up with how we structure the world or how we construe it as being. This kind of empirical evidence about the workings of the mind will thus be indicative of the metaphysics of the world *from the internal perspective*. This is so because, for the internal realist, the notion of a metaphysics only makes sense from the inside, from within the world – the kind of context-free God's Eye view of metaphysics required by Objectivism is devoid of meaning.

Chapter 4
The History of the Psychology of Categorisation.

Introduction.

In this chapter, my aim is to provide a brief and simple account of the history of the psychology of categorisation. I give an account of the classical, probabilistic and explanation-based models of categorisation and an account of the problems associated with similarity-based views. I also show how the explanation-based view differs from and improves upon its predecessors and how it avoids the problems associated with them. It should be borne in mind that this is just a summary of the psychological research and results. I take these results to be correct and as showing what they claim to show. It is in no way my intention to criticise either these particular theories and results or the psychological method in general. In fact, in the previous chapter I explained why I take empirical evidence to be so important for a full understanding of our concepts and of the process of categorisation.

What I *do* want to emphasise (and this is spelt out in the "Evaluation" section of the chapter) is that, in my opinion, the shifts in type of theory which have occurred mark a gradual move away from very tightly defined, all-purpose models to more flexible models. Later models allow for context-sensitivity, relativity and a more human approach to categorisation. They concede that man might actually have a role in shaping the categories of the world and that his concepts may not simply be mirror images of real-world categories.

It can be seen that the classical model of categorisation is an objectivist model and so takes on board all the assumptions associated with that doctrine. The shift away from such tightly-defined accounts towards the more flexible and human-oriented explanation-based account is indicative of the problems and dissatisfactions with Objectivism, as noted in Chapter Two. The explanation-based account, I believe, is consonant with positions such as Hilary Putnam's internal realism and Mark Johnson's experiential realism, where the emphasis is placed on the human agent situated *within* his environment, both being affected by *and affecting* that

environment.

What is Categorisation?

Categorisation plays a crucial role in our conceptual lives. Imagine the number of objects and events which we encounter in a single day, let alone in an entire lifetime. It is vast. It is essential that we have some way of reducing these items to manageable proportions in order that we are in a position to make sense of the world in which we live.

When we categorise, we gather together various distinct items and treat them as equivalent in some way. We selectively store information in memory about these groups of items. This information can then be used as a basis for action, judgement and decision in our future encounters with the furniture of the world.

The role of categorisation is essentially two-fold. Firstly, it results in cognitive economy; it means that we are able to cope with the infinite number of stimuli which we experience. Secondly, it allows us to make inferences about the world in which we live. Objects belonging to the same category will be similar in some sense; what is true of one of these objects may well be true of others in the category. Furthermore, because we can generalise about an item based on the category to which it belongs, we are in a position to infer more about that item than we can ascertain simply by looking at it.

It would seem that categorisation is one of the major uses to which we put our concepts. In very general terms, a category is seen as a class which exists in the real world and the corresponding concept is seen as the mental representation of that category. The psychologists' aim is to look at how we categorise objects that we experience and thus to shed light upon the mental representation of our concepts, on conceptual structure.

The Classical View.

Up until the mid-1970s, a certain view of categorisation was accepted wholesale and without question in psychology. This was known as the classical view and had its roots in Aristotelian philosophy.[93] The key notion here is that of necessary and

[93] For confirmation of this point see, for instance, Smith and Medin, 1981.

sufficient conditions. Each of our concepts has a set of defining conditions (properties) attached to it, which are individually necessary and jointly sufficient for membership in that class. In order to decide of an object whether it is a member of a particular class, all one need do is ascertain whether it satisfies the defining conditions attached to that class. It might, hypothetically, be claimed that the defining conditions for membership in the class of chairs are having a seat, a back, four legs and being made of wood. In this case, an object possessing all four features would qualify as a chair, whilst an object lacking one or more of these features would not be a chair.

It should be fairly obvious that the classical view has some pretty restrictive consequences. It entails that membership in a class is an all-or-none affair. Either something belongs to a class or it does not. There is no question of fuzzy boundaries between classes, with queries as to which category an object falls in to. Objects belong properly in one class only. Each member of the category is as much a member as the next; no provision is made for some objects being perceived as more central or typical members of the category than others. This means that in order to judge that an object belongs to a particular class, all we need is definitional knowledge. That is, all we need at our disposal are the defining conditions which are attached to the class and the ability to decide whether that object exhibits those particular features. General world knowledge (knowledge of how the extension of a concept relates to or interacts with the extensions of other concepts) apparently plays no part in this cognitive process. As was illustrated in Chapter Two, the classical view is a thoroughly objectivist model.

Problems with the Classical View.

During the 1970s, people began to call into question the classical view, which had been taken for granted for so many years. One of the most famous research groups associated with this area was that of the psychologist Eleanor Rosch and her colleagues. The majority of the work produced in opposition to the classical view was experimental in nature.[94]

The major problem with the classical view is that it is very clear that the categories which we use *do* have fuzzy boundaries. Class membership is apparently not an all-or-none affair. It is sometimes unclear whether an object belongs in one category or

[94] It should be noted that this research deals with natural categories (natural and artifactual kinds).

in another. For instance, do rugs and television sets belong to the class of furniture? Is a minneola an orange? Is a hammock a chair or a bed? Is chess a game or a sport? It is possible to think up further examples *ad infinitum*.

McCloskey and Glucksberg, amongst others, built a psychological experiment around this intuition. They asked their subjects to judge whether certain objects belonged to certain familiar categories. These objects had antecedently been judged for degree of typicality in the class by an independent group of subjects. The results show that for those objects which were judged to be of only an intermediate level of typicality (e.g. bookends = furniture), subjects frequently disagreed with one another as to whether the objects belonged in the category or not. Further, after a one month period, it was found that subjects showed a tendency to change their own decisions about category membership. In contrast, with objects which were judged to be highly typical of a category (e.g. chair = furniture) and with those which were totally unrelated to the category (e.g. cucumber = furniture), subjects agreed with each other and were consistent in their own judgements.[95]

These results strongly suggest that categories have fuzzy boundaries and that membership is not all-or-none. If it were all-or-none, there should be no disagreement, either inter- or intra-subject, regarding membership. The boundaries between classes should be clear-cut, so that members and non-members can be distinguished easily without reservation.

A second problem with the classical view is that our categories appear to exhibit what has come to be known as prototypicality, which results in certain typicality effects. That is, it seems that as part of everyday life we quite naturally distinguish between typical and atypical members of a class. For example, said of a bean bag, "That's a funny sort of chair!", or of the merits of apples vs. kiwi fruit, "They are proper fruit. I don't like these new-fangled exotic things which are sold in the shops these days!"

Again, this intuition has been exploited in psychological research. There is overwhelming experimental evidence that subjects will readily judge certain members of a category to be better examples, or more typical of the category than others.[96] This appears to be a perfectly natural task and being asked to judge

[95] McCloskey and Glucksberg, 1978.
[96] See, for instance, Rosch and Mervis, 1975; Rosch, 1973.

typicality produces no surprise or difficulty. Furthermore, this tendency has been shown to have effects in relation to other cognitive tasks. For instance, when subjects are asked to list instances of a concept, the more typical exemplars are listed first; when asked whether an object belongs to a category, the more typical the object is of the category, the quicker a response is elicited from a subject; children tend to learn highly typical exemplars of a category before less typical exemplars and so on.

The final problem with the classical view is that, even after all this time, for most natural concepts, no one has yet come up with the requisite sets of necessary and sufficient defining conditions. No sooner do you come up with a set, than you think of instances of the concept which do not satisfy the conditions. For instance, it is quite likely that a bean bag will fail to meet any of the conditions which you might expect to find attached to the concept 'chair'. It seems unlikely that a tiger which lost its stripes or a leopard which lost its spots would cease be considered a tiger or a leopard by us. Up until very recently, books were only printed on paper, but now they can be read on-line.

Again, there is empirical evidence supporting this claim. For instance, Bellezza asked a number of subjects to define certain commonly used nouns. He then repeated the very same experiment one week later. Results showed that different subjects varied considerably in the features which they used to define these nouns. Furthermore, in the repeat experiment, it was found that individual subjects would define the same nouns differently – some of the features used would have varied between experiments.[97]

This would appear to provide evidence against concepts comprising sets of necessary and sufficient conditions. There is both inter- and intra-subject disagreement concerning definitions of common nouns. If the classical view were correct, this should not be so – we should all, as competent language users, have invariant sets of conditions in our heads which we use at all times in defining concepts or in making decisions about category membership.

Although it seems to be widely the case that we are unable to come up with sets of conditions which are definitive of class membership, this does not in itself deal the death blow to the classical view. Maybe we are just not advanced enough as yet to have worked out what the real defining conditions are, or maybe we never will be.

[97] Bellezza, 1984.

However, the cumulative evidence against the view is fairly damning and has led psychologists to look elsewhere for the roots of our concepts.

The Probabilistic View.

Many psychologists were struck by the apparent prevalence of typicality effects in the conceptual domain and it was this feature that formed the basis for further research in this area. Much of the work must be attributed to Eleanor Rosch and her colleagues. They proposed a new view of conceptual structure known as the probabilistic or family resemblance view.[98]

The terminology "family resemblance" is not coincidental. Psychologists were becoming increasingly aware of the work of Wittgenstein. They saw his notion of family resemblance as a strategy with which it might be possible to by-pass the problems which had plagued the classical view.

The idea here is that our concepts comprise sets of characteristic or typical features which are abstracted across instances, rather than sets of necessary and sufficient conditions. These are mentally represented in the form of an abstract summary representation. There need be no real world instance which corresponds to the representation, hence its abstract quality. This means that although many members of the category will possess all or many of these characteristics, there is no necessary requirement that they possess them all. Other members of the category will possess fewer of these characteristic features. For example, we might say that characteristic features of the concept 'bird' are to have wings and feathers, to fly, to sing and to eat worms and these features will be included in the mental representation of the concept. Robins, thrushes and blackbirds possess all these features. On the other hand, ostriches and turkeys lack the last three features, yet they are still members of the 'bird' category. It is these overlapping chains of shared attributes which hold the members of a category together.

Swapping necessary and sufficient for characteristic features means that the probabilistic view can naturally incorporate typicality effects. Those category members which possess all or many of the characteristic features should be judged by people to be typical members, whereas those which possess few of the features should be judged less typical. In family resemblance terminology, this means that

[98] See, for instance, Rosch and Mervis, 1975.

those members with higher family resemblance to the category (those which possess more of the characteristic features) will be judged more typical than those with lower family resemblance. Using the above example, we are likely to view robins as more typical of the 'bird' category than ostriches.

It is assumed that the features or attributes associated with a category are weighted for saliency. Saliency is calculated in terms of the number of category instances which share that particular attribute. For instance, all birds have wings. Thus wings would be highly salient for the 'bird' category. On the other hand, turkeys make gobbling sounds. Very few other birds possess this feature and so gobbling would have very low saliency for the 'bird' category. In order to compute an object's family resemblance to a category, it is necessary to sum all the weighted attributes that it possesses. A high sum will result in high family resemblance score and vice versa for a low sum. For example, many birds have wings and feathers, fly and sing, so these features are likely to be highly salient for the 'bird' category and will have high saliency weights attached to them. Robins possess all of these attributes. Summing these attributes will result in a high family resemblance score for the robin, which explains why robins are considered typical birds.

Theoretically, a similar computation would need to be carried out in order to ascertain whether a particular object was or was not a member of a category. The weighted attributes which it possessed would be summed. If this sum exceeded a certain threshold which had previously been set for category membership, then it would belong in that category. If it failed to exceed this threshold, then it would belong to a contrasting category.

Rosch and Mervis conducted a series of experiments to confirm the intuition that degree of family resemblance directly reflects typicality. Subjects were asked to list attributes which they considered to be characteristic of each of 120 common objects. Each object belonged to one of six common categories. Each of these items had been scored for typicality in its category in a previous experiment. Each attribute listed for each item falling into a particular category was weighted, according to how many items in the category had been listed by the subjects as possessing that attribute. Every item was then measured for family resemblance in its category by summing the weighted attributes which it had been listed as possessing. It was found that family resemblance scores correlated well with previous typicality ratings for the items. Thus it seems that the more an item has attributes in common with other

members of the category, the more it will be considered a typical member of that category. It seems, then, that the probabilistic, unlike the classical view, can account extremely well for typicality effects.[99]

The probabilistic view is also able to accommodate the notion of fuzzy category boundaries, which proved so much of a problem for the classical view. By definition, membership in a category is not an all-or-none affair, according to the probabilistic view. Some members are judged to be more typical than others and this is a direct reflection of their degree of family resemblance. What this means is that some objects may have a very low family resemblance score in a particular category, since they possess very few attributes in common with other members. It will then be unclear whether they belong to that category or to another contrasting category. For example, is a rug which I hang on the wall furniture, or is it a picture? On the probabilistic view, the boundaries between categories may be fuzzy and unclear, which will naturally result in queries such as these.[100]

The Attachment to Similarity.

The one element shared by the above theories is a fundamental attachment to the notion of similarity. The idea is that when we categorise, we group together objects which we believe to be similar. Similarity is construed as something which is out there in the real world, waiting to be perceived or discovered. It is seen as something absolute and unchanging. Similarity is, in its own right, seen as sufficiently constraining to provide an adequate explanation of the categorisations that we make.

Take the classical view. This utilises the notion of sets of individually necessary and jointly sufficient conditions as determinants of category membership. If an object possesses all these conditions, then it is similar to other objects in that category in virtue of possessing those conditions and so is deemed a member of that category. Similarity of features is considered to be an objective fact of the world and it is this similarity which results in specification of classes. Much the same thing holds of the

[99] Rosch and Mervis, 1975.

[100] Around this time, some psychologists propounded what is known as the exemplar view. This advocated the mental representation of concepts in the form of specific instances or exemplars, rather than as abstract summary representations. Proponents of this view were also aiming to replace the classical view. I am omitting to include a discussion of this view in the main text because it was simply another reaction against the classical view, with much the same aims and origins as the probabilistic view, and so is an unnecessary added confusion when giving a history of the shift from tightly constrained to more flexible theories of categorisation.

probabilistic view. Once again, similarity is the key notion. Objects falling into a category are linked by overlapping chains of similarities. Some members of the set will be more similar than others. Most notably, the central or typical members will share most characteristic features in common, yet even the non-typical members of the class will share features in common with some other members. It is these crisscrossing similarities which make the class cohere and in virtue of which a particular item belongs to a particular class.[101]

However, more recently, many psychologists have begun to doubt that similarity can do the job of explaining all the concepts which we have and all the categorisations which we make.[102] It seems that similarity may be too flexible to explain categorisation; it may not be sufficiently constraining.

The problem is, it is not clear that it makes sense to say that two things are similar *per se*. Similarity is not static in this way. Prior to judging two things to be similar, it is necessary to be aware of *what will count as similar in this context.* It is necessary to define a backdrop against which the notion of similarity can come into play. It may in fact be more correct to say that we perceive two things as similar because we group them together in the same category than to say that those two things are simply similar and by dint of their similarity fall into the same category.

Murphy and Medin argue that potentially, any two objects can be arbitrarily similar or dissimilar. They give the example of a lawnmower and a plum, which are similar in that they both weigh less than 10,000 kg, both did not exist 10,000,000 years ago, both cannot hear well, both can be dropped, both take up space and so on.[103] We could equally make up a list of differences.[104] This is why similarity is too flexible to explain categorisation. We need to drop down to a deeper level of explanation. We need to explain *why* we choose certain attributes over others on which to base our similarity judgements and to explain *in virtue of what* we judge members of a category to be similar. To take another example, a Bedlington Terrier seems to share as many similarities with a lamb as with a Great Dane, yet we would classify it as a dog, along with the Great Dane and in contrast to the lamb.[105] Why is this so? Similarity is not sufficient on its own to account for our categorisation. The

[101] It seems that the probabilistic view can at least partially be interpreted as an objectivist model, since it treats similarity as a fixed and static fact of the world.
[102] See Murphy and Medin, 1985, for one of the earliest expressions of this doubt.
[103] Murphy and Medin, 1985.
[104] See Goodman, 1972, for an expression of parallel views about similarity from a philosopher.
[105] This example is taken from Komatsu, 1992, p. 505.

explanation lies a stage deeper, with principles which specify which similarities count and which do not in drawing the boundaries between classes.

This kind of argumentation sheds quite a new light on both the classical and the probabilistic views. Asking subjects to list properties which they consider to be either defining or characteristic of particular concepts is really explanation kicking in a stage too late. The attribute lists that those subjects provide will be biased in a fundamental way. Any attribute list will presuppose some sort of heuristic. These properties have already been judged to be relevant according to that heuristic. It is these heuristics or principles which really have explanatory power and not the attributes in terms of which similarity judgements are made. The heuristic underlies and explains both choice of attributes and judgements of similarity and typicality.

As Douglas Medin puts it:

> "The general point is that attempts to describe category structure in terms of similarity will prove useful only to the extent that one specifies which principles determine what is to count as a relevant property and which principles determine the importance of particular properties. It is important to realise that the explanatory work is being done by the principles which specify these constraints rather than the general notion of similarity. In that sense similarity is more like a dependent variable than an independent variable."[106]

Rips' Experiment.

Rips conducted an influential experiment in order to illustrate that judgements of similarity may differ systematically from judgements of categorisation.

Subjects were presented with stories concerning members of both natural and artifactual kinds which underwent certain changes. Some of these changes Rips termed as 'essential' changes and others he termed as 'accidental' changes.[107]

In one story, an animal was described as starting off in life with characteristics which

[106] Medin, 1989, p.1474.

[107] It is important to note that Rips is not talking about metaphysically essential and accidental properties. Psychology deals with epistemology and not metaphysics. Rips is simply aiming to distinguish changes which *we might consider fundamental* from those which we might not consider fundamental in relation to categorisation judgements and this is irrespective of whether or not there are such things as metaphysically essential properties.

are typically associated with birds and as gradually maturing and taking on characteristics typically associated with insects. Rips considered this an essential change, presumably some kind of change determined by internal genetic structure. In another story, an animal with characteristics generally associated with birds was described as eating vegetation contaminated by hazardous waste. Following this incident, its behavioural and physical characteristics were described as gradually becoming like those generally associated with insects. Rips considered this an accidental change, since the alteration only involved surface characteristics and was unlikely to alter the animal's genetic structure. It is important to note that the words 'bird' and 'insect' were never used during the experiment. Neither were the phrases 'accidental change' or 'essential change' mentioned. Thus, the subjects were given no idea of the ways in which they were expected to view the story.

In the case of artefacts, essential changes involved change in designer's intended function and accidental changes involved changes in the artefacts' physical characteristics, such as colour, which did not affect function. So, in one story, it was explained that something fitting the description and function of an umbrella was re-decorated so that it looked much more like something fitting the description of a lampshade, although it was still used to shelter its owner from the rain. In another story, an object was outlined which fitted the description of an umbrella. Subjects were then told that the designer designed this object with an intended function which fitted that of a lampshade. Again, the words 'lampshade' and umbrella' were not used during the experiment – subjects were simply given detailed descriptions of the imaginary objects.

The results of the experiment showed that in the case of both natural kinds and artefacts, accidental changes had a much greater effect on similarity judgements than on categorisation judgements, whilst essential changes had a much greater effect on categorisation judgements than on similarity judgements. So, the animal which had undergone an accidental change tended to be categorised as a bird by the subjects, even though they judged it to be dissimilar to other birds and not at all typical of a bird. Conversely, the animal which underwent an essential change was, in its bird-like stage, considered to be very similar to and very typical of birds, but was not considered likely to be a bird.[108] Likewise, the umbrella which underwent the

[108] Keil, 1989, conducted a similar experiment whereby subjects were told about a raccoon which had undergone surgery in order to make it look and behave as a skunk would. The subjects judged that despite these superficial changes, this animal was still a raccoon and not a skunk.

accidental changes which made it look more like a lampshade was still categorised as an umbrella by subjects, despite bearing high similarity to lampshades. The object which was intended for use as a lampshade by the designer was judged to be a lampshade, despite the fact that it bore high surface similarity to umbrellas. Subjects insisted that even if it were put to use in the future as an umbrella, it would still be a lampshade, since that was the intended function of the designer.[109]

These results indicate that similarity does not lie at the bottom of our categorisation. Similarity and categorisation judgements can vary independently of one another and so are not inextricably linked. It looks as though there are deeper explanatory principles which constrain and can overrule our notions of similarity in the domain of categorisation.

Other Problems with Similarity-Based Views.

Categorisation as Attribute Matching.

Another feature of the early views on categorisation is that they treat our concepts as being mentally represented in the form of features. The classical view characterises our concepts as corresponding to sets of necessary and sufficient conditions, which take the form of features or properties. The probabilistic view treats our concepts as comprising groups of characteristic features or properties which category members are likely to possess.

The problem is that is beginning to look less and less likely that our concepts are *only* represented by means of feature lists. I have already made the point that similarity alone is not constraining enough to explain categorisation – we need to ask why we pick certain features as relevant and why we count certain things as similar. The question now becomes, what holds these feature lists together and why do we isolate the features that we do? Once again, we need a deeper and more sophisticated level of explanation.

If you look closely, it seems that concepts comprise much more than simple feature lists. We might say that a typical bird has wings and feathers, flies, sings and eats worms, but this really tells us very little – it is not a true representation of our knowledge, it is just an enumeration of properties. What is missing is the vast quantity of *relational* knowledge which we have and make use of all the time in

[109] Rips, 1989.

cognitive activity. We assume that the reason birds tend to exhibit these properties is that they are related by some kind of underlying, genetic structure and it is this that gives the list of properties its sense and brings it to life. So, it seems that our concepts might involve the representation of some kind of relation between features.

It also seems likely that our conceptual knowledge involves some notion of relations between concepts themselves. A concept does not just have life and meaning in virtue of itself. The world is an essentially relational place and things tend to have relevance in relation to other things. For instance, a chair may become significant only in relation to a number of things. It is made to be sat upon and so people who might want to sit down immediately come into the picture. Some chairs are used for sitting at tables – we recognise the association between these two sorts of object. Others are made for comfort and they tend to be found in people's living rooms, in front of the television and adorned with cushions. Or take our concept of a dog. We know that dogs are often pets, owned by people, they tend to live in houses and they like to eat meat. Our concept of the dog involves a lot more than a list of dog-like properties – it involves knowledge of how the dog operates in the world and how it is related to other things in that world. So, concepts are not isolated entities – it looks as though a large proportion of their meaning may be made up from their interaction with other concepts. We are moving from a notion of definitional conceptual knowledge to one of encyclopaedic conceptual knowledge.

Murphy and Medin give a neat example to illustrate that categorisation might involve much more than simple attribute-matching. We might categorise someone who jumps into a swimming pool with all their clothes on at a party as intoxicated, even though swimming fully clothed is not an attribute which comes into our concept of 'drunk' and even though this is not something that we have ever seen someone who is drunk do before. Murphy and Medin argue that this shows that categorisation as attribute matching may just be too limited. It may be that at least some categorisations are in fact much more akin to inference processes. We make use of our general knowledge of how people who are drunk are apt to behave and are then able to apply this non-specific knowledge to an instance of a kind which we have never before experienced.[110]

[110] Murphy and Medin, 1985, p. 295.

The Importance of Context.

The early views on categorisation allow little or no room for differences in context. The classical view involves fixed lists of necessary and sufficient conditions whilst the probabilistic view involves lists of characteristic features, which tend to be pretty fixed, except that category members will possess more or less of them, depending on their centrality in the category or on their typicality.

However, it seems that which attributes are relevant to a particular concept at a particular time will vary with context – not all attributes are appropriate all of the time. For instance, one thing which all banknotes have in common is that they are flammable. Under normal circumstances this is unlikely to be a particularly relevant attribute, but when you are deciding what to grab on your way out of your burning home, their flammability may become very important. In that context, the attribute becomes highly salient.[111] Similarly, our concept of a grand piano may include the attributes of being heavy and of being tuneful when played. When I move house, the fact that the piano is very heavy to lift upstairs becomes highly relevant to the removals company but its tunefulness is likely to be uninteresting to them. Yet, when the piano tuner visits my house, tunefulness is of key importance, whilst weight is irrelevant.[112]

It looks, then, as if our concepts may be much more flexible than early accounts have been willing to allow. Apparently, the importance or relevance of attributes will vary with the occasion of use.

Instability of Conceptual Representation.

Finally, it is worth mentioning that different people are likely to have slightly different representations for the same concept. This is because different people have different personal experiences and this is likely to shape the way they look at the world. For instance, someone who has recently been bitten by a dog is likely to have a slightly different concept of a dog from someone who has not recently been bitten. It is likely that viciousness will play a key role in the bitten person's concept 'dog', whilst it may play a very small, if any, role in the concept of the non-bitten person.[113] Again, this is something not allowed for by early views of conceptual representation. Differences in personal experience are not seen to enter the picture at all. A concept

[111] This example comes from Murphy and Medin, 1985, pp.299-300.
[112] This is based on evidence from Barclay et al., 1974, cited in Barsalou, 1989.
[113] This example is adapted from material in Barsalou, 1989.

is seen as something static which is shared by all. If someone has a concept which differs from the norm it must mean that their concept is in some way incorrect, not simply that their experience has been different.

People in different cultures may well have concepts which vary to an even greater degree than those between people in the same culture. For instance, for people living in India, the concept 'cow' may include the information that a cow is not a source of meat, due to the extremely important religious significance that the cow has. However, a British person's concept of a cow is likely to incorporate the idea of a common food source, whilst religion is unlikely to play any part in the concept at all. Likewise, a French person may see a frog as something which is eaten on special occasions. A British person is less likely to see the frog as a food source.[114]

Once again, it seems that the early views on categorisation are simply not flexible enough to deal with this kind of variation. They treat concepts as fixed and uniform, yet it is clear that they *can* vary, both between cultures and even between people in the same culture.[115]

The Explanation-Based View.

In recent years, due to all the problems associated with similarity-based views of categorisation, there has been a move towards what are called explanation-based theories of categorisation.

It is quite difficult to give a unified account of the explanation-based view, since it is still in its infancy and theories are still in the process of being developed. However, it is possible to make some general remarks about the motivations and aims of the various theories.

First and foremost, the explanation-based view is an essentially *relational* account. What this means is that relations are explicitly included in our conceptual representations. It is these underlying relations which cause our concepts to be coherent – they make sense of the categories which we form. Relations between properties which we tend to associate with instances of a concept are therefore made

[114] These examples are adapted from material in Barsalou, 1989.

[115] It should be noted that there are other problems associated specifically with the probabilistic view. However, these are somewhat irrelevant to my aim of producing a history of the development of views of conceptual representation from static to more flexible. See Komatsu, 1992 for a summary of these other problems.

explicit. So too are the relations between different concepts which serve to give these concepts their life and meaning. The relations which are specified between the properties associated with a concept tend to be causal or functional. For instance, we might think that the reason that most birds have wings, feathers, beaks etc. is that these features are the product of some kind of underlying genetic structure which is the cause of the properties and so provides an explanation for why birds are grouped together. In the case of chairs, we might think that the reason that many of them have a seat and a back is that they are made for people to sit on. Those particular features mean that chairs can fulfil the function of being sat upon. This functional relationship serves to explain why we group chairs together.

The relations which are specified between concepts serve to give them their life. So, our concept of gold may include the knowledge that it is an element which is mined, that it is stored in the vaults of the Bank of England and that wedding rings tend to be made of it. Gold is meaningful for us by dint of its relation to many other concepts which play a part in our lives. Likewise, our concept of a hammer may include thoughts about tool boxes full of tools, of picture hooks needing to be hammered into walls and of our healthy hatred of DIY. The important point is that our concepts are not just specifications of one particular kind of object, rather in order to have a concept of anything at all, you need to be acquainted with the roles that that kind of object plays in the world in relation to other things and to people. If this is not specified, you will be left with a highly attenuated or impoverished notion of what a concept is.

It can be seen that the explanation-based account is concerned with knowledge which we have of the world surrounding us. It recognises that concepts are not isolated or self-contained. It recognises that we play a part in shaping our concepts. Things are not necessarily categorised because of the way they are in themselves, but because of the role they play in our lives. Our concepts might not be a simple mirror-image of the world's categories – they may be something much more cognitively interesting than that.[116]

[116] There are a number of pieces of empirical evidence which go towards illustrating that it is likely that our concepts are underlain and driven by theories and explanations, but it is beyond the scope of this chapter to discuss these. See Komatsu, 1992 and Medin, 1989 for summaries of some of the evidence.

How the Explanation-Based View Copes with Problems for Similarity-Based Views.

The explanation-based account does not do away with similarity completely, but it takes much more of a back seat than in previous views. The point is that the functional or causal underlying relationships are what cause us to say that two things are similar. They provide principles which serve to constrain and make sense of our category groupings. It may be that straightforward perceptual similarity works in some cases as an explanation – a thrush and a robin look pretty similar, it would seem – but what *really* explains this grouping is some kind of underlying relationship between the two instances. With many cases of categorisation, apparent straightforward similarity will just not do. To take the earlier example of the person jumping into a pool fully clothed being labelled as 'intoxicated', we might reason here that excessive drinking tends to alter people's perception of things and to make them behave in an unusual manner and do things that they would not normally do. This is the underlying causal explanation of why the person jumped into the pool and so also the reason why we categorise them as intoxicated.

Of course, the explanation-based view deals with much more than attribute lists. Although properties will inevitably play a major role in any conceptual representation, relations between these properties are very important too. Categorisation becomes much more than mere attribute-matching. It involves knowledge of why certain attributes are treated as highly salient by us, of what underlies the choice of these over and above others. The attributes which are salient are taken to be indicative of certain underlying relationships which we use to govern our decisions concerning category membership.

Since the explanation-based view is much less constrained than earlier accounts, changes in our concepts due to changes in context need not pose a problem. With a change in context, different attributes of the concept become relevant because that concept is playing a different role in relation to other concepts. So, in the example of the fire and the banknotes, 'flammability' becomes highly salient because banknotes are made of paper and so burn very easily. They are also valuable and so you might decide that they would be a good thing to pick up in a fire. Under normal circumstances, however, 'flammability' is not a salient attribute of banknotes, since their function is purely economic and they have little relation to burning houses. With the change in context, a slight shift in underling relations occurs, which

accounts for the change in saliency which is attached to certain properties. Since the explanation-based view is concerned with these relations, it can, unlike earlier views, happily accommodate this kind of shift.

As a result of the explanation-based view relating instances to concepts not on the basis of similarity between characteristic or defining features, but on the basis of unifying underlying explanations, instabilities in conceptual representation can be naturally incorporated. Concepts which differ between countries or cultures may be unified by slightly different explanations. In France, frogs' legs are a national delicacy and so a Frenchman's concept of a frog should have a strong connection with food and perhaps restaurants. On the other hand, due to differing culinary traditions in Britain, a British person's concept of a frog is less likely to be related to food. At the level of individuals, different experiences may result in slightly different concepts. This will be due to the fact that different conceptual relations will be salient for different people. Someone who has recently been bitten by a dog is likely to see viciousness as a key feature of dogs, whereas someone who has only ever encountered friendly dogs, may not consider viciousness to be a feature which is highly associated with dogs. These two people have had very different experiences with dogs, they see them as behaving in different ways and this may result in two slightly different concepts. Once again, different relations are salient for these people as a result of the differences in their personal experience. Concepts need not be static and invariant according to explanation-based accounts, but may naturally differ between people.

Psychological Essentialism: an Explanation-Based Model.

I think it is useful to look at one particular explanation-based model in more detail. This helps us to see how some of the very general remarks about explanation-based theories cash out in terms of more explicit and concrete psychological models.[117] The particular model I am going to look at is called psychological essentialism and is put forward by Douglas Medin and Andrew Ortony.[118]

One of the major tenets of psychological essentialism is recognition of the psychological fact that people's mental representations of things in the world reflect

[117] There are, in fact, various models which follow an explanation-based structure, but in the interests of brevity, I am going to look at only one candidate. See Komatsu, 1992, for a summary of other models.

[118] Medin and Ortony, 1989.

the belief that these things have essences.[119] It is imperative, however, to realise two things. Firstly, Medin and Ortony are talking about people's *representations* of things, not the things themselves. This makes sense, since concepts are mental representations of things. Secondly, (and as a consequence of the first point) they are not advocating metaphysical essentialism. In fact they dismiss metaphysical essentialism as a "logically implausible doctrine".[120] The point is that irrespective of whether or not things have essences, people seem to *believe* that they do.

Medin and Ortony draw attention to Rips' experiments, in which subjects were unwilling to agree that changes in superficial surface properties of objects indicate a change in category membership. This suggests that the subjects believed that the objects possessed some kind of deeper properties or essence in virtue of which they were those particular objects and that this essence persisted through surface changes.[121] Secondly, Medin and Ortony feel that people are likely to endorse essentialism, since most scientific enquiry is aimed at uncovering the underlying true nature of phenomena, rather than just describing their observable properties.

Medin and Ortony feel that the problem with earlier accounts of categorisation is that they very much deal only with surface properties of objects which tend to be perceptual in nature. The early accounts then judge similarity in terms of these surface properties. However, Medin and Ortony feel that this is not adequate to explain how we categorise. For instance, we sometimes judge things which are not perceptually similar to be in the same category. A whale looks much more like most fish than most mammals, yet we judge it to be a mammal. Furthermore, it is not clear that family resemblance structure categories can be made to cohere on the basis of surface similarity alone. A stool, a bean bag, a jumbo cushion and a swivel chair might all belong in the 'chair' category, yet it is not clear that they are perceptually similar enough that the category could be said to form a cohesive group.

Their solution is to make a link between the surface and the deeper properties

[119] McNamara and Sternberg, 1983 provide experimental evidence for the fact that people tend to believe in necessary and sufficient defining conditions, even though they are not always able to say what these conditions might be.

[120] Medin and Ortony, 1989, p.183.

[121] The same point can be made with regard to Keil's experiment involving the skunk/racoon. And, more recently, Malt (1990) reports an experiment in which subjects were told of a plant halfway between a marigold and a dandelion in appearance. They judged that it would make more sense to say, "We'd have to ask an expert to tell us which it is" than to say, "I guess you can call it whichever you want", suggesting that they felt that this plant had some sort of hidden essence in virtue of which it would fall into either one group or the other, despite its apparent ambiguity.

possessed by the members of a category. The deeper properties form the essence for psychological essentialism and it is these deeper properties which really determine category membership and on which similarity judgements will be based. However, there is a very strong non-arbitrary link between the deeper and the surface properties, since the deeper ones actually constrain and in some cases generate the surface ones. So, for instance, people might assume that the 'real' criteria for being an aeroplane impose fairly strong constraints on the surface features of aeroplanes, such as size of wings, shape of body etc. Likewise, people might assume that some kind of underlying genetic bird structure will result in possession of wings, beaks and feathers. In other words, the psychological essence of an object imposes some kind of limitation upon the surface properties which that object possesses.

Rather than drawing an arbitrary line between surface and deep properties, Medin and Ortony suggest that the attributes associated with a concept lie on a continuum of accessibility, stretching from highly accessible, through to relatively inaccessible and hidden. At the very deepest level, these attributes will fill what Medin and Ortony call an "essence placeholder" in our conceptual representations. This placeholder may be filled with different kinds of thing, depending on the concept in question. It may be filled with a list of defining conditions which are seen to make an object an instance of the category or with a set of beliefs which is more akin to a theory of what it is to be an instance than to an attribute list.[122] It might also contain the belief that there are experts who really know what makes a thing the thing that it is and so what qualifies an object to be an instance of a particular concept, even if laymen cannot specify what this might be. The beliefs which fill this placeholder will constitute the essence of psychological essentialism and it is these which are held to constrain surface properties.

It seems, however, that on an everyday basis we resort to surface perceptual similarity in order to classify objects and generally this works well for us. We would classify a large feline animal with stripes as a tiger because it *looks like* other tigers. Surface similarity appears to be a good categorisation heuristic. Medin and Ortony suggest that this is because surface properties are constrained by deeper properties. Similarity of surface properties is normally a good bet for similarity of deeper

[122] For instance, the case of labelling the person who jumps into the swimming pool fully clothed as 'intoxicated'. Here, you are making a kind of inference. You have a general theory of how someone who is drunk might behave and by using this theory you conclude that the most likely explanation for this person's behaviour is that he is drunk.

properties. Appearances do not normally deceive us. However, occasionally, as in the whale/mammal example, the heuristic breaks down and one can then see clearly the crucial role that deeper properties play in the categorisation process.

We can now see how psychological essentialism qualifies as an explanation-based account of categorisation. According to Medin and Ortony, people adopt what might be called an "essentialist heuristic". That is, they act as if things in the world possessed essences which make them the things that they are. However, in many cases people are not able to say what this essence is – all they are acquainted with are the surface perceptual features of things. As a result, they make judgements about similarity and category membership on the basis of these surface features and most of the time this is successful and yields judgements which correspond to the psychological essence. Yet the reason this works is that the essence constrains or generates the surface properties which things have. The deep and surface properties are inextricably linked. This functions as a theory or explanation of how the world works, which enables people to make predictions and inferences about things in the world. If deeper properties constrain surface ones, then on the whole it will be a good bet that things that look alike tend to share the same deep essential properties (the ones which really count in terms of theories about category membership).

It is, however, still extremely important to remember that Medin and Ortony are not talking about how the world is in itself but only about how we represent the world as being. They are not advocating metaphysical essentialism at all. Their point is simply that people behave as if things had essences and that this is simply one way of coping with the multitude of objects in the world. It is just one way in which people can legitimately treat a number of objects as equivalent in some respect, in that they are all treated as having the same psychological essence. This provides a theory or explanation as to why these particular objects might be categorised together.

Evaluation of Shifts in Theories of Categorisation.

It seems to me that there is a clear distinction between the earlier theories and later explanation-based theories of categorisation. There is clearly a shift from a very tightly defined notion of what our concepts are and what categorisation involves to a much more flexible, less formal approach. There is also a clear shift from the notion of man passively reflecting the real categories of the world by his concepts to the notion that man makes an active contribution to those concepts and categories.

Take the earliest classical view. Here the key word is precision and the aim is the most cognitive economy possible. Our concepts are said to be defined in terms of sets of necessary and sufficient conditions (properties), which constitute the essence of what it is to be an instance of that concept. This means that in order to be an instance of the concept, an object must, necessarily, exhibit all the defining properties. If it does not, it is not an instance of that concept. There is no question of something being more or less a member of a category or class, there is no question of any query concerning the class to which an object belongs.

Categorisation is a very simple and precise affair, according to the classical view. All that is required is that we know the defining conditions attached to each concept. It is then just a case of comparing an object's properties against these conditions in order to decide whether it is an instance of that particular concept. There is no room for any variation in our concepts.

The classical view treats our concepts as simple mirror images of the categories which are assumed to exist in the natural world. This means that every natural category has a set of ascertainable conditions attached to it which determine membership in that category. These conditions are invariant and really do exist out there in the world. All *we* have to do is discover them. Once we have done this, we should have a concept, consisting of a set of defining conditions, corresponding to each category. Thus the ultimate aim of the classical view is to produce a conceptual taxonomy which directly corresponds to the real taxonomy which is believed to exist in the world. This taxonomy is ultimate, correct and just waiting to be discovered. The structure of our concepts is then an unadulterated reflection of the underlying world structure. The reason we are unable to come up with sets of necessary and sufficient conditions for many of our concepts is simply that we do not, as yet, have sufficient knowledge of the requisite metaphysics. In other words, the classical view is *the* objectivist model.

What this means in very general terms is that man makes no real contribution to categorisation. The only work he has to do is to discover the requisite sets of defining conditions. His role is an essentially passive, receptive one. Once he discovers the world's true structure, this will be reflected in his concepts. In fact, a very harsh line is drawn between man himself and the environment in which he operates. Man is not seen as a part of that environment or as having any role in shaping it or the categories which make it up. He is set apart from the world and his role is to reveal and reflect

the world as it is, in itself.

The probabilistic view is the beginning of a process of recognition of the faults of the classical view. By this time, there was a vast amount of psychological evidence which seemed to show that our natural concepts are not governed by necessary and sufficient conditions, that category membership is not an all-or-none affair. We find it entirely natural to judge particular objects as better or worse instances of a particular category, whilst some members of some categories seem to possess very few properties in common – certainly nothing amounting to a set of defining conditions. The probabilistic view directly incorporated these typicality effects by replacing the notion of a set of defining conditions, all of which must be exhibited by all category members, with the notion of a set of characteristic features which would be exhibited by the most typical members, but which need not all be exhibited by all members. This is certainly taking a large step away from the classical view, but it is a change concerning *conceptual structure only*. The probabilistic view recognises that an all-or-none defining structure is unlikely to be applicable to many of our concepts, but it pays no real attention to man's role in the environment. The human agent is still assumed to have a passive role in relation to that environment and the categories which it is assumed to inherently comprise.

The probabilistic view still tends to treat our concepts as reflections of the world in itself. The concepts which we have will correspond to categories which are out there in the world to be discovered by anyone who cares to look. Rosch tells us, "A working assumption of the research on basic objects is that (1) in the perceived world, information-rich bundles of perceptual and functional attributes occur that form natural discontinuities, and that (2) basic cuts in categorisation are made at these discontinuities...basic objects are the categories that best mirror the correlational structure of the environment."[123] Mervis and Pani explain, "Because all co-occurrence patterns of attributes are not equally probable, the attribute structure of the world may be used as a reasonable basis for the assignment of objects to particular categories."[124]

The point is that certain attributes in the natural world have a higher probability of co-occurring that others. For instance, a beak and wings are much more likely to co-occur with feathers than with fur. The assumption is that those attributes which co-

[123] Rosch, 1978, p.31.
[124] Mervis and Pani, 1980, p.497.

occur most frequently form natural categories – categories which exist in the real world and which our concepts should mirror. Thus, wings, beaks and feathers frequently co-occur and so form the natural category which we call 'birds'. However, no attempt is made to explain in what way or why we relate these groups of features with one another. No mention is made of underlying theories or explanations which might account for the salience of those particular attributes.

The other problem with both the classical and probabilistic views (as explained previously), is that they both use an absolute notion of similarity in order to explain how we categorise. However, it seems that this notion of similarity simply cannot explain enough – we need some kind of deeper theoretical principles in order to explain why we judge certain things to be similar, why we categorise certain things together and why we pick certain attributes as salient for the similarity/categorisation judgements that we make.

I see the similarity problem as just a further extension of treating man as a passive agent in the categorisation process and of treating his concepts as reflecting a unique category structure which is inherent in the world. The early views treat similarity as something static and inflexible. Things in the world are seen to be similar *per se* and this similarity is just there waiting to be discovered. Similarity is not seen as being dependent on certain explanatory theories or principles which specify which among the many similarities are to count. Man is not assigned any kind of instrumental or active role. He simply perceives similarity, but plays no part in picking out or choosing between the numbers of similarities which can be seen to run through the natural world.

Explanation-based theories begin to adopt a very different approach to categorisation from that adopted by the earlier classical and probabilistic views. It seems that it is finally being acknowledged that concepts might be more than passive reflections of categories which already exist in the natural world. It is acknowledged that man plays some part in shaping both his concepts and the categories which he perceives in the world. As Medin puts it:

> "It is tempting to think of categories as existing in the world and of concepts as corresponding to mental representations of them, but this analysis is misleading. It is misleading because concepts need not have real-world counterparts (e.g. unicorns) and because people may impose

> rather than discover structure in the world. I believe that questions about the nature of categories may be psychological questions as much as metaphysical questions...The world could be partitioned in a limitless variety of ways, yet people find only a minuscule subset of possible classifications to be meaningful. Part of the answer to the categorisation question likely does depend on the nature of the world, but part also surely depends on the nature of the organism and its goals. Dolphins have no use for psychodiagnostic categories."[125]

It can be inferred from explanation-based accounts that man's contribution to the categorisation process takes the form of the construction of theories or explanations. These theories or explanations then act as deep principles which underlie our judgements of similarity and category membership. For instance, we judge all tigers to be similar and to belong together in one class, yet, by theory, we judge that they look similar because they all have similar underlying genetic structure – perhaps they all possess the same kind of DNA. Our concepts are much more than simple reflections of categories in the world. They involve fairly complex theories which serve to explain why we judge things to be similar in a certain respect. The theories can be seen to govern the similarities rather than vice versa.

I have also emphasised that explanation-based theories are essentially *relational* accounts. This means that concepts do not have meaning in isolation. Rather, they gain their meaning from their association with other concepts or with other things in the world. On the one hand, the features which are typically associated with certain concepts are seen to cohere because of certain causal or functional structures which we posit in explaining the concepts/categories which we use. On the other hand, we typically relate concepts to other concepts in our mental representations. So, our concept of a tiger might involve representations of cats, of the jungle, of ferocious animals, of stripes, of endangered species. Once again, man himself can be seen as having an active role in the categorisation process. It is only through our interaction with the world and our recognition of certain things as salient that we relate certain concepts to other concepts. Particular connections or relations are salient for us and these then become part of our conceptual representation. For instance, the average British person is likely to associate tigers with zoos and to keep away from them unless they are behind bars. On the other hand, imagine a people who practise an

[125] Medin, 1989, p.1469.

initiation ceremony into manhood. This ceremony involves hunting down a tiger, sustaining a bite from it and surviving the bite. The survival marks the transition from boyhood to manhood. Such a person's concept of a tiger is likely to be somewhat different from our own. He will relate the tiger to particular events, traditions and ideas which are pertinent for him, but which are unlikely to be pertinent for us. It is in this kind of way that man has a hand in structuring his concepts together with the categories which he recognises in the world.

Very generally, the move from the earlier views on categorisation to the later explanation-based views marks a shift from conceptual structure involving definitional knowledge to conceptual structure involving encyclopaedic knowledge. This means that our concepts are now seen as involving a much broader base of general world knowledge than was formally supposed. Our concepts apparently do not consist (only) of clear and distinct lists of attributes which together specify the meaning of the concept. Rather, they involve knowledge of relations between concepts or objects and of relations underlying attributes which tend to be associated with particular objects or concepts. This suggests a much more holistic approach. Much of the encyclopaedic knowledge, as I have begun to show, is not just a reflection of the way the world is, but involves human beings in theory construction and in certain biased ways of looking at the world. The environment is no longer seen as a self-contained place and the division between man and the world in itself has begun to blur.

Chapter 5
Philosophy Attacking the Psychology of Categorisation.

Introduction.

In this chapter, I want to look at an attack which is launched on the psychology of categorisation by the philosopher Georges Rey. He focuses his attack on the book *Categories and Concepts* which is written by the psychologists Edward Smith and Douglas Medin. This book has come to be considered the classic review of all the early theories of categorisation. It therefore covers the classical, probabilistic and exemplar views together with combinations of these, plus the problems concerning these views. The book was published in 1981 and so predates the birth of the explanation-based views. The major conclusion which Smith and Medin reach is that the findings of psychological studies of categorisation do not fit the classical view of concepts. Rey's major aims are to question that the psychological findings have any relevance for a theory of concepts and to attempt to shore up the classical view of concepts.

Rey's response to Smith and Medin is important within the context of this thesis, since he is the only *philosopher* of whom I am aware who has written directly on the subject of the psychology of categorisation. Moreover, Rey is an *objectivist* philosopher. He is therefore of the opinion that 'correct' categorisation involves reflecting the world's metaphysical structure and that our psychology or epistemology is irrelevant for telling us anything about the world's metaphysical categories. He believes that the psychology of categorisation can tell us about the way in which we access our concepts but can tell us nothing about concepts themselves (since they are simply reflective of metaphysics). Rey's attack on the psychology of categorisation therefore provides us with a concrete instance of Objectivism, as characterised in Chapter Two of this thesis, coming into conflict with more epistemological or human-oriented accounts.

In the first few sections of this chapter, I give an account of the debate between Rey

and the psychologists. I then give an analysis of this debate in terms of the line which both parties draw between metaphysics and epistemology. What we see is that both parties concur in the separation of metaphysics from epistemology, placing psychology firmly in the domain of epistemology. Importantly, both seem to agree that the scientist, when making a classification, is concerned *solely with metaphysics*. It is this position, together with the objectivist assumptions underlying it, that this thesis aims to challenge.

The Debate.

The Core and Identification Procedure.

At this stage, it will be useful to give a brief account of a certain psychological theory of categorisation which is considered in *Categories and Concepts*, since much of the debate centres around this theory.[126]

This theory involves making a distinction between the core and identification procedure of a concept. A concept's core will contain deep diagnostic properties whilst the identification procedure contains properties which are commonly used to categorise real world objects. When making decisions about concept membership, people will rely mainly on the properties relating to the identification procedure. These tend to be perceptual in nature, easily accessible, but not always foolproof. The core properties will be much more reliable, but they are less accessible and therefore less available for people when making decisions about concept membership. To quote Smith, Medin and Rips:

> "Specifically, identification properties are often useful for a 'quick and dirty' categorisation of objects, and such properties tend to be salient and easy to compute though not perfectly diagnostic of concept membership; core properties, on the other hand, are more diagnostic of concept membership, but they tend to be relatively hidden and hence less accessible for rapid categorisation."[127][128]

[126] This theory was not discussed in Chapter Four of this thesis. This was so for two reasons. Firstly, it adds nothing to a general overview of the history of theories of categorisation; it is simply an attempt to combine a classical view of concepts with the probabilistic psychological data. Secondly, as a theory, it won little support from the psychological community as a whole.

[127] Smith, Medin and Rips, 1984, p.267.

[128] The core and identification procedure can, very clearly, be seen as a forerunner of the explanation-based model of psychological essentialism.

Take the example of gender. Here, identification properties may include things like clothing, hairstyle and voice, whilst core properties might involve possession of sexual organs. Or take gold. Identification properties might include its colour, the uses to which it is put and where it is mined, whilst core properties might involve being a basic element with atomic number 79.

The basic idea behind this kind of theorising is two-fold. Firstly, it is a means of combining the traditional classical view with the typicality effects associated with the probabilistic view. The idea is that the core of a concept might have a classical type structure, perhaps being composed of essential, defining conditions. On the other hand, the identification procedure may have a probabilistic structure. That is, the properties which we use in making decisions about concept membership will only be characteristic of the class. Not all possible members of the class need possess these properties and so members will be considered more or less typical, dependent on property possession. Those which possess a large number of properties exhibited by other members will be judged highly typical and vice versa for those possessing few properties exhibited by other members. Secondly, the core properties can be seen as a means of backing up or justifying rapid judgements made on the basis of identification properties alone. For example, we might come across some substance which looks very like gold, but on closer inspection we may find that it is not gold, since it is not a basic element with atomic number 79 – it does not possess the requisite core properties.

As with the model of psychological essentialism discussed in the last chapter, it is important to bear in mind that psychologists deal with epistemology only. Therefore, the distinction between the core and identification procedure does not mark a distinction between metaphysics and epistemology. The core properties are simply to be seen as properties which are less accessible but more reliable than identification properties in making epistemological decisions about concept membership. This is so regardless of whether there are such things as metaphysical core properties. As Smith, Medin and Rips put it:

> "...we hold that the distinction between the identification procedure and core of a concept has nothing to do with metaphysics. Rather, it is a distinction between two kinds of properties that may be used in epistemological categorisation...our distinction centres on notions like salience, computability and diagnosticity, not on metaphysics...with

> regard to [core properties] we are trying to characterise the layperson's theory of the nature of things...[not] to characterise the nature of things..."[129]

Rey's Attack.

Rey launches his attack in his paper, "Concepts and Stereotypes". I will run through this paper, summarising the arguments and points which are put forward. It is helpful to bear in mind here that Smith and Medin came to the conclusion, in the light of the psychological evidence, that the classical view is not adequate to account for many of our (natural) concepts and that the probabilistic view presents a more plausible theory of conceptual structure. Rey is therefore essentially attacking the probabilistic view and the validity of the probabilistic evidence for a theory of concepts.

The first major point which Rey wants to drive home is his belief that the psychology of categorisation continuously confuses questions of metaphysics with questions of epistemology. More precisely, he thinks it confuses questions of conceptual *identity* with ones of conceptual *access*. He believes that the psychology gives us interesting results concerning conceptual access, but that this tells us nothing about concepts themselves, or their identity conditions, or what conditions must hold in order for us to be able to say that someone *has* a particular concept.

Rey goes on to draw a sharp line between metaphysics and epistemology or between, "issues surrounding *how the world is* (what exists, what is true) and issues surrounding *how we know, believe, infer how the world is*."[130] He argues that concepts themselves play a metaphysical role in that they are able to provide the basis for certain metaphysical claims. To use Rey's own example, the concept 'cow' can provide the basis for the claim that Elsie is a cow by specifying certain facts about Elsie in virtue of which she is a cow. He believes that, "Concepts in this role may be regarded *in isolation*, as providing the principles of classification for what is to count as their instances..."[131] So, there are metaphysically real categories and corresponding essential or defining properties, according to Rey – the world comes with its own unique structure. On the other hand, there is also an epistemological role played by concepts. This involves how we *tell* whether something is an instance of a concept. Rey uses the example of gender. There are metaphysical conditions in virtue

[129] Smith, Medin and Rips, 1984, pp.267-268.
[130] Rey, 1983, p.243.
[131] Rey, 1983, p.243.

of which someone is either male or female, but in order to tell what sex someone is, we rely on certain extraneous features such as hairstyle, mode of dress and pitch of voice. These are not features in virtue of which someone is either male or female, rather they are perceptually salient features which we are able to use to come to a reasonable decision concerning category membership. Rey's contention is that the psychology of categorisation tells us much about the latter epistemological role of concepts, but nothing at all about what, in a later paper, he dubs the "only serious notion"[132] of concepts – their metaphysical role.

Rey continues this line of argumentation, claiming that the word 'categorising' can refer either to how things are *correctly* categorised, or to how people actually engage in the process of categorising (i.e. it can refer either to matters of metaphysics or to matters of epistemology). He suggests that Smith and Medin themselves consider this very distinction when they look at the possibility of a concept core distinct from its identification procedure. He then concludes that they fail to take this distinction seriously and that in fact they see no reason to distinguish between the metaphysical and epistemological roles that concepts play. Rey, however, considers this to be a very important psychological distinction. In particular, if we fail to make this kind of distinction, then different people who use different identification procedures for the same concept or one individual who uses different identification procedures for the same concept at different times will have to be considered as having different concepts. Thus you and I might use slightly different perceptually salient features to decide that a robin is a bird, but surely we would not want to say that you and I have different 'bird' concepts. Furthermore, if no distinction is made, any change in belief about how to tell whether or not something is an instance of a concept will have to count as a change in concept. So, Rey is concerned that the epistemological aspect of concepts and hence the discipline of psychology is neither able to account for conceptual identity nor conceptual stability – only concepts in their metaphysical role can account for these phenomena, by providing a list of defining conditions in virtue of which something is the thing that it is. What makes two things instances of the same concept is that they conform to the same definition. What constitutes two people having the same concept is that that concept has one and the same definition attached to it, despite the fact that these people may associate different identification procedures with it.

[132] Rey, 1985, p.298.

The second major point in Rey's paper involves a defence of the classical view of concepts. This defence turns on what Rey calls, "an entirely implicit assumption attached by Smith and Medin to the Classical View".[133] This is the assumption that in order to have a particular concept, one must *know* its defining conditions. Rey, however, taking his cue from the work of Putnam and Kripke, argues that whether anyone knows that a concept has defining conditions is entirely irrelevant to whether or not it *does* have defining conditions. It might be that concepts have necessary and sufficient defining conditions and that these are true descriptions of real world classes or categories, yet this has nothing whatsoever to do with what we believe or infer to be true of the world. It might be that we do not know these conditions now, or that we will never know them. The point is that we can quite competently make use of a concept without knowing its defining conditions. According to Rey, the division between metaphysics and epistemology is absolute.

Rey posits what he calls the "Hypothesis of External Definitions". This asserts that, "...the correct definition of a concept is provided by the optimal account of it, which need not be known by the concept's competent users."[134] He tells us that by "optimal account" he means, "an account of a concept, all relevant issues and possible evidence considered, than which no account (metaphysically) could be better."[135] Although there are probably scientific experts who can provide definitions for many of our ordinary language terms, there need not be. Definitions are divorced entirely from anyone's ability to provide them. There will always be certain conditions in virtue of which something is the thing that it is, quite independent of anyone knowing what those conditions are.[136]

Rey believes that his hypothesis effectively protects the classical view from all the psychological evidence which has been adduced against it. By maintaining an unerring division between metaphysics and epistemology, what people know or believe is entirely irrelevant to what concepts actually represent – the real structure and true classes of the world. Psychology, according to Rey, has nothing to contribute to metaphysics, it only has relevance for epistemology, so whilst the evidence it has

[133] Rey, 1983, p.251.
[134] Rey, 1983, p.255.
[135] Rey, 1983, p.255.
[136] Rey notes that he considers Putnam's hypothesis of the division of linguistic labour to be too epistemological, since it rests on the notion that there are scientific experts who are able to provide us with the requisite definitions. His own hypothesis is therefore entirely metaphysical and is in no way dependent on human knowledge.

adduced is highly relevant for epistemological issues of conceptual access, it leaves the metaphysical issues (such as conceptual identity and stability) untouched and a haven remains in which the classical view is able to survive, fit and well.

It is clear from the above account that Rey is an objectivist, in the tradition of Putnam and Kripke. In particular, he supports objectivist metaphysics by indicating that the world possesses it own unique category structure determined by essential properties which is quite unaffected by human cognitive activity. He is also a supporter of objectivist epistemology, since he talks of correct categorisation, which involves a simple reflection of the world's inherent category structure. Furthermore, he makes it very clear that the definitions which are attached to metaphysical categories are simply a matter of the way the world is and have nothing to do with our knowledge, beliefs or understanding of those definitions or of the world. He sees any account of how human beings come to grasp the meaning of concepts or of how their concepts relate to the world as unnecessary. For these reasons, he takes psychology to be irrelevant for matters of metaphysical categorisation.

The Reply to Rey.

Edward Smith, Douglas Medin and Lance Rips team together to reply to Rey's attack in their paper, "A Psychological Approach to Concepts: Comments on Rey's "Concepts and Stereotypes"".

Firstly, they deal with Rey's dual criticism that metaphysics and epistemology is conflated throughout Smith and Medin and that psychology is only capable of shedding light on the epistemological role of concepts. Smith, Medin and Rips point out that they are in total agreement with Rey that in order to provide identity conditions for concepts, you need to concern yourself with metaphysics. This is because despite the fact that different people might make use of slightly different identification procedures, we would still want to say that they have the same concept. Furthermore, the properties which we use in identifying instances of a concept are not necessarily tied to that concept and were these properties to change, we would not want to say that the concept itself had changed. Thus it must be metaphysical facts which act as identity conditions for concepts.

The other reason for Smith, Medin and Rips' agreement is that, as they quite rightly state, they never claimed to deal with anything other than the epistemology of categorisation in Smith and Medin, contrary to Rey's belief that they drew a

distinction between metaphysics and epistemology and then chose to ignore it. They go on to argue that the reason Rey was misled was that he failed to understand the distinction between the core and identification procedure discussed in Smith and Medin. Again, they are quite right. Rey assumes that the core of a concept involves metaphysics and that the identification procedure of a concept involves epistemology or how we decide something is an instance of a given concept. However, as I made clear earlier, neither the core nor the identification procedure has anything to do with metaphysics – both deal with epistemology.

Smith, Medin and Rips maintain that the difference between Rey and themselves is that Rey, by stressing the need for metaphysically defining conditions, is attempting to characterise the nature of things. This involves the notion of scientific experts who are attempting to specify what these defining conditions are. They, as psychologists, however, are, by talking about core properties, trying to characterise the layperson's theory of the nature of things. This involves mental representation and epistemology only. To quote Smith, Medin and Rips, "...we concur with Rey [about]...the need for metaphysics in formulating identity conditions of concepts, but we do not think there are any claims about metaphysics in Smith and Medin or in most other psychological studies of categorisation that we are familiar with."[137]

Secondly, Smith, Medin and Rips deal with Rey's criticism that epistemological categorisation cannot account for either inter- or intrapersonal conceptual stability. They agree with Rey that if stability of concepts is to be equated with sameness of concepts, then epistemology will not be able to account for stability, since same concepts require identity conditions, hence metaphysics. However, they claim there is another kind of stability which they call "communality" which can be equated with similarity of mental contents. So, if two people or one person at different times have similar mental contents, then conceptual stability will ensue. This kind of stability can be accounted for in terms of epistemology.

Smith, Medin and Rips go on to point out that many psychological experiments have required subjects to list attributes which they believe to be characteristic of a particular class of objects. The results show that at least some of the attributes listed are mentioned by the majority of subjects in each sample and these common attributes appear to be the same in different samples and in different experiments.[138]

[137] Smith, Medin and Rips, 1984, p.268.
[138] They cite Malt and Smith, 1982, Rosch and Mervis, 1975 and Smith et al., 1984 in support of this

It is also likely, they argue, that core properties provide another source of communality – we all tend to share very basic core beliefs, such as that instances of a particular animal share some kind of genetic structure, or that offspring must be of the same biological type as their parents. Furthermore, interchange with other people results in communality – if you judge an object to be an instance of a particular concept by dint of certain identification and core properties and enough people disagree with you, you are likely to revise the properties which you used in order to come to that particular decision. The result will be increased interpersonal communality. Smith, Medin and Rips feel that there is considerable evidence for the kind of communality of concepts that they are advocating and they feel that for everyday comprehension and the normal purposes of communication, this is quite enough to secure conceptual stability without the need to invoke sameness of concepts and so metaphysics.

Rey's Reply to Smith, Medin and Rips.

Rey gives a very short reply to Smith, Medin and Rips in his paper, "Concepts and Conceptions: a Reply to Smith, Medin and Rips".

The major claim in this paper is that metaphysics *is* of relevance to psychology. Rey says, "The metaphysical – to my mind the only serious – notion of concept has consequences not only for the usual philosophical and scientific questions of what *is* what, but also for questions about what people *believe* to be what. People, when they classify things, are as much engaged in metaphysics as the philosophers and scientists who are paid to do so, which tends to make the "layperson's theor[ies] of the nature of things" (Smith, Medin and Rips, p.5) continuous with those of the experts. We should therefore expect the layperson's concepts sometimes to coincide with the expert's, despite many of the differences in beliefs and conceptions that invariably arise. In at least this way, the metaphysical notion of concept has implications for psychology."[139]

In other words, how both experts and laypeople classify things in the world will be heavily influenced by metaphysics, but (from Rey's 1983 paper) the way in which we go about classifying things in the world (epistemology) can tell us nothing whatsoever about metaphysics. The world is as it is and nothing can alter it. Metaphysics, for Rey, is supreme.

---

claim.

[139] Rey, 1985, pp.297-298.

Analysis.

I now want to make some general comments about the work which has been summarised so far in this chapter and treat this as an entry point into questions which will be debated throughout the rest of this thesis.

The Blurring of the Line.

> "Concepts...would seem...to be about the world and how *it* divides up; not about how we might divide up our methods of investigating it. The interest is in the ducks themselves, not in our ways of knowing them."[140]

The above quotation signals Rey's determination to draw an unfaltering and absolute line between metaphysics and epistemology, with metaphysics playing the superior and determining role. My aim in this thesis is to question, in the case of categorisation and concepts, the rigidity of the line which has been drawn by Rey and others. I suggest that this line is actually blurred and that our concepts are much more than mirror images of categories which exist in the real world.[141] In particular, I will go on to argue that *even at the level of the expert scientist, classification involves a mixture of metaphysics and epistemology.*[142] It simply does not make sense to divorce the way the world is from the ways in which we access the way the world is.

Interestingly, it seems that at this time, the psychologists themselves, when challenged, concurred in the drawing of a very clear line between metaphysics and epistemology, placing themselves, as psychologists, squarely in the domain of epistemology. Smith, Medin and Rips therefore agree with Rey that a metaphysical notion of concepts is needed in order to provide identity conditions for concepts. They also agree that if stability of concepts is used to mean sameness of concepts, then the metaphysical notion of a concept is again needed to account for this stability. They therefore seem to accept that the notion of a totally predetermined and fixed metaphysics is a notion that makes sense and so that only metaphysics can provide the key to *correct* classification of objects in the world.

Despite their apparent concurrence with Rey on the above points, Smith, Medin and Rips' theorising almost seems to contradict this agreement. They entertain the notion of a distinction between the core and identification procedure of a concept, assuring

[140] Rey, 1983, p.250.
[141] These issues have already been touched on in Chapters Two, Three and Four of this thesis.
[142] The bulk of the material in support of this claim appears in Chapters Six and Seven.

us that core properties have nothing to do with metaphysics. Rather, they represent a different set of (epistemological) properties which are used to back up decisions about category membership made on the basis of identification procedure properties. It almost looks as if Smith, Medin and Rips are claiming that epistemology *does* have relevance for metaphysics, in the sense that what grounds our decisions regarding category membership is not simply the way the world is in itself, but rather the theories which we have about that world. It looks as though they are tending towards the idea that our concepts are not simple reflections of real world categories, but that we, as human beings, have a certain degree of influence (by means of our theories) over category formation and over which properties are to be taken as salient for matters of category membership.

However, Smith, Medin and Rips throw away this possibility when they go on to say that the difference between them and Rey is that Rey is trying to characterise the nature of things (his domain being metaphysics) and so he speaks in terms of the special sciences, whilst they are trying to characterise the layperson's theory of the nature of things (their domain being epistemology) and so they speak in terms of lay science. This suggests that they agree with Rey's assertions that things in the world do have a certain predetermined nature and that the role of the special sciences is to discover the metaphysically true facts in virtue of which things are what they are. They seem to be suggesting that only non-experts have theories about the way the world is. Expert scientists, however, are in a different position. They do not deal in theories, so much as search for a correct classification of the objects in the world, which involves a reflection of the world's inherent structure, untainted by theory or any other human factor. It looks, then, as if Smith, Medin and Rips are endorsing Rey's strict metaphysics/epistemology distinction regarding *scientific* categorisation and are accepting that metaphysics remains unaffected by epistemology.

The question I want to explore, however, is *whether it is right to claim that at the scientific level, 'correct' classification is just a matter of reflecting metaphysical structure*. In the last chapter, we saw that particularly with the advent of explanation-based accounts of categorisation, there has been a distinct blurring of the line between metaphysics and epistemology with Douglas Medin, for instance, telling us that, "...people may impose rather than discover structure in the world..." and that, "...questions about the nature of categories may be psychological questions as much

as metaphysical questions..."[143] It seems as if some psychologists are now becoming aware of the importance of the role that epistemology plays and so are beginning to realise that perhaps psychology has something to offer metaphysics after all.

My suggestion is that this blurring of the line between metaphysics and epistemology is as much a facet of classification by scientific experts as it is a facet of classification by non-experts. As a result of this, I will go on to argue, in Chapters Six and Eight, that classification by scientific experts fits an explanation-based model. This means that scientific definitions of terms or concepts are not simple reflections of the structure and properties of the natural world, but they actually incorporate theories or explanations which highlight certain properties as salient for matters of category membership, whilst downplaying other properties. A consequence of this is that it is possible to arrive at competing classifications of the natural world which are equally objective and equally scientific. These competing classifications will be based on different explanations which highlight different properties.

The above line of argument suggests that it is misguided to speak of the world as possessing a unique structure and to assume that metaphysics is the ultimate arbiter in scientific determination of the categories of the world. It suggests that the line between metaphysics and epistemology really is much less clear-cut than Rey and other like-minded objectivist philosophers would have us believe.

Conclusions.

In this chapter, we have looked at the debate between the philosopher Georges Rey and various psychologists of categorisation. This debate is important for two reasons. Firstly, Rey is the only philosopher who has discussed the psychological evidence regarding concepts and categorisation. Secondly, Rey's attack is objectivist in nature. It therefore provides us with an example of the kind of objectivist presuppositions which I discussed in Chapter Two in action. This is useful, since it is those very presuppositions which I seek to undermine in the course of this thesis.

We have also seen that, at least prior to explanation-based views of categorisation, psychologists themselves have supported the notion that psychological evidence can only inform us about epistemology and not about metaphysics or the nature of things.

[143] Medin, 1989, p.1469.

They accept that the world as structured uniquely and independently of human cognition is a coherent notion. According to this picture, the expert scientist is allocated the role of discovering the world's structure. This is a task which involves metaphysics alone and has nothing whatsoever to do with epistemology.

In the next chapter, I begin to break down this kind of picture of the expert scientist and of scientific classification. I do this by means of three case studies which examine how and according to what principles scientists actually classify entities in the natural world.

Chapter 6
Case Studies.

> "Natural science does not simply describe and explain nature, it is part of the interplay between nature and ourselves."[144]

Introduction.

In this chapter, my aim is to adduce a large body of evidence against the objectivist interpretation of scientific classification, as outlined in Chapter Two. That is, I aim to challenge the objectivist assumption that the world is divided uniquely and inherently into a number of natural kinds and that the role of science is to uncover and reflect that inherent taxonomy. Challenging this assumption also involves challenging a number of cognate assumptions, such as that essentialism is true, that categorisation is simply about the way the world is, that epistemology is irrelevant for matters of categorisation and therefore that the only role which human beings have is that of passively reflecting the metaphysics. I further argue that the evidence adduced in this chapter strongly supports the contention, which I mentioned in Chapters One and Five, that the explanation-based account of categorisation is a suitable model of expert scientific as well as of lay classification.

Perhaps the most important insight which the evidence in this chapter provides us with is that the scientist is forced to make a choice or decision regarding which properties or patterns of similarities are to count as relevant for matters of classification. This is because the natural world does not come with a unique preferred description – it has different patterns of similarity and regularity running through it which can count as salient for the purposes of classification.[145]

This does not mean that there is no such thing as an external world or that we, as human beings, construct our own world. Of course scientists are constrained in their division of the natural world by the ways in which that world is and when they make

[144] Werner Heisenberg quoted in *The Penguin Dictionary of Modern Quotations*.

[145] This claim bears a resemblance to Dupré's promiscuous realism (Dupré 1981;1993), although both my initial suppositions and the consequences which I draw from this claim are very different. Dupré's work is discussed in more detail later in this chapter.

a choice as to which properties are salient for classification, they are, most definitely, choosing between real and objective properties. For instance, it might be argued that scientists have chosen to take atomic number as the principle of classification of the chemical elements.[146] In making this decision, however, they are bound by certain objective facts. They are bound by the fact that each chemical element *does* have a certain number of protons in the nucleus of the atom – a certain atomic number. If scientists were to claim that gold has atomic number 78, they would be wrong, because gold has atomic number *79* – it has 79 protons in the nucleus of the atom and not 78.

Using Case Studies.

The evidence adduced in this chapter against the objectivist interpretation of scientific classification is empirical. Firstly, I review the variety of species concepts which have been current in recent biology. What we discover from this is that there is a debate raging over which is *the* correct species concept for matters of biological classification. Proponents of one concept aim to show how their concept is universally correct, whilst illustrating the shortcomings of the other competitors. One recent response to this debate has been to argue for a radical species pluralism whereby the complexity of the biological world is acknowledged and it is suggested that one single species concept could not do justice to the extent of that complexity. Rather it is suggested that different species concepts reflect different aspects of biological reality with some fitting particular purposes and particular organisms better than others. Hence to try and choose one concept over and above all the others is in fact misguided.

My review of the species problem in recent biology plays a very important role in lending support to my project in the second half of this chapter, which is an analysis of three case studies taken from the history and sociology of science. Firstly, my review of recent work shows us that the debate concerning species is just as fierce and contentious now as it was in the history of taxonomy, hence it is not possible to dismiss the conclusions which I draw from the case studies as applicable only to the *history* of science – they are equally applicable to work *currently* going on in science. In particular, the recent debate concerning species concepts can be seen as the modern extension of the competition between accounts of classification based on

[146] This is in fact argued in Chapter Seven.

morphology and accounts of classification based on biosystematics which is catalogued in John Dean's case study. Secondly, my review of arguments for species pluralism in relation to contemporary biology lend support to my own claims concerning the richness of the natural world and the variety of ways in which it is possible to classify the entities within it.

There are a number of reasons for examining case studies in connection with these issues. Firstly, the doctrine of Objectivism or metaphysical realism is essentially *theoretical*. That is, it represents a philosophical conception of how the world (metaphysically) is and it represents a philosophical conception of either how science is or how it ought to be. Such theories, I submit, are not adequate and do not stand alone. We also need to examine what is *actually* the case in science and what science *actually* tells us about the way the world is in order to test the strength and suitability of our theories. There is a wonderful quotation from David Hull which aptly sums up my opinions on this issue:

> "It in instructive to note that during the extensive discussion of the applicability of Kripke's notion of a rigid designator to such terms as "tiger", no one saw fit to see how those scientists most intimately concerned actually designated tiger...that no one bothered tells us something about the foundations of conceptual analysis."[147]

Of course, the objectivist might simply counter this with the claim, à la Rey, that what scientists or what *any* human beings do is simply irrelevant to metaphysics. However, many objectivists, such as Kripke and the early Putnam, cite the findings of science in support of their claims. To backtrack and claim that science is irrelevant to metaphysics when it suits you to do so is not good enough. Furthermore, one of the notions which has been emphasised throughout this thesis is that, in the domain of categorisation, epistemology does have relevance for metaphysics. This is made particularly explicit in Chapters Two and Three, where the merits of internal/ experiential realism and their relation to categorisation and the use of empirical evidence are discussed.

Secondly, science has an extremely high profile in our society. People tend to accept the scientist's word and assume that he has the definitive answers. Furthermore, I think people tend to assume that in the scientific domain there is only one available

[147] Hull, 1984, p.638.

answer and this is the one which reflects the way the world is. I suspect that it is these kinds of sentiment which lead to a widespread acceptance of essentialism and Objectivism regarding natural kinds and scientific classification.[148] The following studies serve, however, to overthrow this kind of conception of science, in particular of scientific classification. We learn, instead, that there is not always one fixed answer, that scientists adhere to different theories and produce competing classifications.[149]

Modern Concepts of Species and Biological Taxonomy.

There are a large number of different species concepts mentioned in recent literature. However, David Hull tells us that by 1971 it was possible to isolate three quite distinct schools of systematics – phenetics (or numerical taxonomy), evolutionary systematics and cladistics.[150] I outline and discuss each of these schools and their competing beliefs and aims. I also take a look at what could be termed a further (very recent) 'school' – pluralism. Adherents to this 'school' claim that there is not one uniquely correct all purpose form of classification, but that different kinds of classification are able to coexist and serve different purposes.[151]

Phenetics.

Phenetics represents a school of morphologically based taxonomy which grew up in the 1960s and 1970s. This school is most commonly associated with the work of Sokal and Sneath and was originally centred on the University of Kansas at Lawrence.[152]

The basic notion behind phenetics is that biological organisms should be classified according to *overall similarity* (incorporating similarity of function, form and biological role) which is to be calculated using numerous characters. Furthermore, each character used in calculating overall similarity is to be given equal weight – no

[148] We saw, in Chapter Four, that subjects in psychological experiments tend to assume that natural entities have essences and that scientists are in a position to state what those essences are.

[149] My account of these case studies will, of necessity, be very brief. I will concentrate on aspects which are relevant to my thesis and by so-doing, much of the original author's detail will inevitably and unfortunately be missed out.

[150] Hull, 1988, p.129.

[151] There is a debate which often runs alongside that of biological species concepts. This concerns the question of whether species are sets (spatio-temporally unrestricted) or individuals (spatio-temporally restricted). However, each debate is self-contained, hence I will not be discussing the kinds/individuals question in this thesis.

[152] Hull, 1988, Chapter 4.

character is to be preferred over and above any other. This work is mathematical in nature – each character is first recorded in numerical form and then overall similarity is calculated by algorithmically manipulating the characters with the aid of computers to produce a diagrammatic chart of the relationship between the entities to be classified, known as a phenogram. This diagram of phenetic distances can then be converted into a biological classification. Pheneticists adopted the name 'numerical taxonomists' for themselves in order to reflect this mathematical bias.

The aim of phenetics is to produce uniform and objective classifications and adherents to this school believe that such classifications can be produced by means of explicitly formulated statistical methods, as discussed in the previous paragraph. Pheneticists are of the opinion that overall similarities which are there for one and all to see must be taken as prior to a priori speculation concerning biological process. As David Hull reports:

> "Classifications should unequivocally represent the resemblances exhibited by organisms without regard to their descent. Descent could then be indicated in an accompanying diagram."[153]

Phenetics can actually be construed as a revival (in more respectable and more modern form) of traditional notions of taxonomy which claim that classifications should be based on morphological discontinuities between entities which are discerned by perception alone. Later on in this chapter, I discuss such traditional notions of taxonomy in more detail when I look at John Dean's case study taken from the history of botany.

Despite the enthusiasm and commitment of its original proponents, phenetics has largely fallen into disfavour. For instance, John Dupré tells us that:

> "...in large areas of biology where more theoretically favoured approaches seem impracticable or inapplicable – especially microbiology and to a considerable extent botany – morphological conceptions of taxonomy remain more respectable. It is, however, quite widely accepted that some independent criteria for assessing the relative significance of morphological features are required. This seems necessary on philosophical grounds to avoid the difficulties with making sense of absolute similarity, and on biological grounds to assure that properties

[153] Hull, 1988, p.119.

selected have suitable evolutionary or other theoretical significance."[154]

I have already discussed some of the philosophical problems with the notion of similarity in Chapter Four in connection with the psychological explanation based view of categorisation. During the next couple of sections of this chapter, we will see how important the theme of evolution has become for taxonomy.

Dupré does go on to stress, however, that morphology can be interpreted in such a way that a more useful and meaningful taxonomy might be based upon it. I will discuss this further in the section headed "Pluralistic Species Concepts".

The Biological Species Concept.

The biological species concept is typically associated with the work of Ernst Mayr. Under this concept, species are defined as:

> "Groups of actually or potentially interbreeding populations which are reproductively isolated from other such groups."[155]

The idea here is that the members of a species together comprise a reproductive community and so respond to one another as potential mates. The species also comprises an ecological unit which interacts as a whole with other species occupying the same ecological area. Finally, the species forms a genetic unit which comprises a sizeable interacting gene pool, in contrast to an individual member of the species which "is merely a temporary vessel holding a small portion of the contents of the gene pool for a short period of time."[156]

Mayr goes on to emphasise that a species is a *protected* gene pool which has so-called "isolating mechanisms" which protect it from the potentially harmful flow of genes from other pools. He explains that genes from the same pool combine harmoniously because they have become adapted to one another as a result of natural selection, whilst the mixing of genes from different pools will result in disharmonious combinations, hence mechanisms which prevent such mixing are favoured by natural selection.

It is clear, then, that the biological species concept (unlike phenetics) rests upon the presupposition of an inextricable link between taxonomy and evolution and Mayr in

[154] Dupré, 1993, p.45.
[155] Mayr, 1963, p.19 (quoted in Sokal and Crovello, 1992).
[156] Mayr , 1984, p.533.

fact states that:

> "An understanding of the nature of species, then, is an indispensable prerequisite for the understanding of the evolutionary process."[157]

Once again, the biological species concept may be identified with what John Dean calls experimental taxonomy or biosystematics in his case study taken from the history of botany, which I discuss later in this chapter.

Limitations of the Biological Species Concept.

There are a number of limitations of or problems with the biological species concept, some of which are articulated by Mayr himself.

Firstly, the majority of species comprise aggregates of numerous local populations which exchange genes. However, the further away populations are from one another, the more likely it is that their respective characteristics will differ. Mayr lists the differences that can occur – species which are widespread may have terminal populations which do not interbreed even though contiguous populations in the chain do interbreed; populations may become reproductively isolated whilst remaining morphologically very similar; populations may diverge morphologically yet continue to interbreed; populations may quite happily interbreed yet, when their *habitats* change, they may become reproductively isolated; isolating mechanisms are built up slowly over time and so can be imperfect and incomplete; isolating mechanisms are established at different rates in different populations, hence it is possible that in some areas two populations are reproductively isolated, whilst in others, they interbreed.[158]

Although these problem cases are simply a result of the very gradual nature of speciation, the question remains, how are we to determine what is and what is not a species in the sense of the biological species concept. There appears to be no concrete answer to this question and Mayr notes:

> Determination of species status of a given population is difficult or arbitrary in many of these cases."[159]

---

[157] Mayr, 1984, p.531.

[158] Mishler and Donoghue (1992) list similar problem cases for the biological species concept. They also note that whilst proponents of the concept consider that gene flow maintains the cohesion of the species, there are other factors which may also confer cohesion, such as internal homeostasis or the state of the external environment. They therefore conclude that it is not possible to give a univocal statement concerning cohesion and its causes.

[159] Mayr, 1984, p.537.

Secondly, by its very nature, the biological species concept cannot be applied to asexual organisms – interbreeding between populations has no relevance here. It is therefore necessary to use some other set of criteria to determine species membership in such cases. Mayr suggests that:

> "...there seem to be rather well-defined discontinuities among most kinds of uniparentally reproducing organisms. These discontinuities are apparently produced by natural selection from the various mutations that occur in the asexual lines (clones). It is customary to utilize the existence of such discontinuities and the amount of morphological difference between them to delimit species among uniparentally reproducing types."[160]

Thirdly, Sokal and Crovello (1992) claim that the biological species concept is subject to what they term "phenetic bottlenecks". By this they mean that, due to time limitations, all such systematists will initially be forced to group individuals and populations on the basis of phenetic similarities and not on whether they do or do not interbreed, thereby cutting down on the number of questions which actually need to be asked regarding interbreeding. Thus, whilst the definition of the biological species concept makes no mention of phenetics, its determination will, of necessity, always do so. Even where crossing tests have been performed, they claim that the basic species definition is still phenetic since any consequent statements will be based on *phenetic* inferences from the small number of crosses actually performed. They therefore conclude that the biological species concept is not a necessary part of evolutionary taxonomy, since most of the evidence for evolutionary taxonomy is based on phenetics and not interbreeding.

> "We are left with what is essentially a phenetic criterion of homogeneous groups that show definite aspects of geographic connectedness and in which we have any evidence at all on interbreeding in only a minuscule proportion of cases."[161]

Although Sokal and Crovello's argument may sound convincing at first blush, it does not stand up to analysis. Their claim is that the biological species concept is not a relevant part of evolutionary taxonomy, since most of the relevant evidence is based

[160] Mayr, 1984, p.538.
[161] Sokal and Crovello, 1992, p.46.

on phenetics and not interbreeding. However, the point they miss here is that the only reason evolution or interbreeding is salient at all, is that we are focussing on the *biological* species concept – and it is the *biological* species concept which aims to reflect elements of the process of evolution by concentrating on interbreeding of individuals and populations. If you abandon the biological species concept, you abandon along with it your interest in evolutionary taxonomy (unless you replace it with another species concept which invokes an *evolutionary* based taxonomy – which phenetics does not). Sokal and Crovello's mistake is to confound the principle behind the biological species concept (which is the reflection of the link between evolution, interbreeding and species) with the process by which that principle is achieved (which may involve the use of phenetic similarities for speed and ease). Their claim that we are left with an essentially phenetic criterion (see previous quotation) similarly misses the point, since what (essentially) motivates a phenetic criterion is the desire to produce a classification based on all observable properties possessed by an organism where *all* these properties are given *equal* weight, whereas what (essentially) motivates the biological species concept is a desire to reflect patterns of interbreeding between organisms. What we are left with is therefore anything but an *essentially* phenetic criterion.

Phylogenetic Species Concepts.

There are a number of variants which can be listed under the heading of 'phylogenetic species concepts' and David Hull has in fact commented on this variation.[162] There also seems to be some ambiguity concerning terminology – sometimes such positions are termed 'phylogenetic' and other times they are termed 'cladistic'. Hull notes that it was Mayr who originally coined the term 'cladism' to refer to phylogeny[163] and the implication seems to be that the two terms refer (at least roughly) to the same (group of) theories.

Despite all this variation, the central tenet underlying (earlier accounts of) phylogenetic classification is that classifications should reflect genealogy or evolutionary branching patterns. In order for a group of organisms to form a species, they must share some kind of common ancestry. There is therefore a direct link between phylogenetic taxonomy and the history of evolution.

"Phylogenetic definitions are thus firmly rooted in the concept of

[162] For example, Hull, 1988, Chapter 7 and Hull, 1989, Chapter 10.
[163] Hull, 1988, Chapter 4.

evolution, that is, of common descent...*What is both necessary and sufficient is being descended from a particular ancestor.*"[164]

Willi Hennig is perhaps the best known proponent of cladism. He was of the opinion that traditional hierarchical classifications were not complex enough to reflect all the details of phylogenetic development. He therefore opted for one particular element in phylogeny – the sister-group relationship. Two taxa (A and B) will represent a sister group if they are more closely related to one another that they are to any other taxon (C), the proximity of relationship being based on characters which A and B share with one another but do not share with C. The sister-group relationship is collateral (not ancestor-descendant), hence A and B must share a more recent common ancestor with one another that either one of them does with C. None of the taxa within the statement of a sister-group relationship are said to be ancestral to any other. Despite that fact that Hennig recognised two types of phylogenetic relationship (the sister-group and the ancestor-descendant relationships), he still insisted that a truly phylogenetic classification concerns itself with sister-group relationships alone.

There are two more important aspects of Hennig's account of cladism. Firstly, the sister-group relationship can be represented in a phylogenetic diagram (or cladogram), in which speciation events are represented as the splitting of a single line (the stem species) into two (the daughter species which form a sister-group). Hennig argued that when the ancestral species splits, it must be considered extinct and only when this splitting occurs should new species be recognised.[165] Thus an ancestral species cannot exist alongside its descendants and so two concurrently existing species can only be connected by the sister-group relation, never by the ancestor-descendant relation.

Secondly, Hennig believed that taxa should be monophyletic. This ordinarily means that a taxon should include only the ancestral species and its descendants, but Hennig extended this, arguing that a taxon should include the ancestral species together with *all* its descendants. This limitation has some surprising consequences, as illustrated by the following examples borrowed from Elliott Sober.[166] Crocodiles and lizards retain many of the features of their common ancestor and so are highly similar but

[164] De Queiroz, 1992, p.300.

[165] It should be noted that G.G. Simpson argues that speciation occurs when a descendant population becomes reproductively isolated from its ancestors and not when an ancestral population gives rise to two descendant populations which are reproductively isolated from one another. (Cited in Kitcher, 1984.)

[166] Sober, 1993.

crocodiles and birds have a more recent common ancestor which is not an ancestor of the lizards. A cladistic classification would therefore group the crocodiles and birds together since they form a genealogical unit apart from the lizards. Secondly, placental and marsupial wolves independently evolved a number of similar characteristics. However, placental wolves actually form a genealogical unit with moles. This means that the cladist would group the moles and placental wolves together on the basis of their evolutionary branching patterns.

With regard to the above rather surprising examples, John Dupré notes that, "Probably a majority of those sympathetic to phylogenetic taxonomy are in fact committed only to the much weaker demand that classification not be inconsistent with the genealogical tree."[167] He explains that this more conservative attitude will generally require that all the members of a taxon be monophyletic but will not necessarily require that the taxon include the ancestral species together with *all* its descendants. A more conservative position would therefore classify crocodiles and lizards separately (as has traditionally been the case), motivated by the belief that taxonomy should, to a certain extent, reflect similarity and difference independent of phylogeny. Furthermore, the more conservative approach allows for the possibility of speciation *without* splitting (anagenetic speciation), where enough change within an undivided lineage might seem to warrant demarcation of more than one species.

One such weaker version of cladistics is the position known as pattern cladism. The people associated with this view are Gareth Nelson and Norman Platnick. According to pattern cladists, "...species are simply the smallest detected samples of self-perpetuating organisms that have unique sets of characters."[168] They abandon Hennig's claim that separate species can only be distinguished as a result of speciation events and instead claim that cladistics is concerned with "detectable changes".[169] Nelson and Platnick seem to be trying to develop a much more general account of cladism than the one provided by Hennig, arguing that terms such as 'cladism' and 'cladogram' "...unfortunately were intended to have an explicitly evolutionary significance pertaining to the actual branching or speciation events of phylogeny."[170] and instead stating that under their account, "*Cladograms* depict structural elements of knowledge."[171][172]

[167] Dupré, 1993, p.48.
[168] Nelson and Platnick, 1981, p.12 (quoted in Hull, 1988, p.248).
[169] Nelson and Platnick, 1981, p.35 (quoted in Hull, 1988, p.248).
[170] Nelson and Platnick, 1981, p.139 (quoted in Hull, 1989, p.152).
[171] Nelson and Platnick, 1981, p.14 (quoted in Hull, 1989, p.152).

Phylogenetic Species Concepts: the Pros and Cons.

Different biologists hold differing opinions regarding the merits (or otherwise) of the phylogenetic species concept.

Joel Cracraft (1992) lists a number of ways in which he believes a phylogenetic species concept to be more adequate than a biological species concept. Firstly, with the phylogenetic species concept, taxonomic units are, by definition, equivalent to known evolutionary units, which leads to a greater clarification of evolutionary issues. Secondly, the phylogenetic species concept clarifies the distinction between recognising species taxa and analysing geographic variation. Thirdly, the phylogenetic species concept makes it clear that subspecies cannot have ontological status as evolutionary units. Fourthly, the phylogenetic species concept does not use data about reproductive isolation to demarcate species. This means that even in cases where two sister-taxa hybridise, each can still be held to be a separate species. Fifthly, the phylogenetic species concept focuses more attention on the geographical history of the species that does the biological species concept. Sixthly, the phylogenetic species concept places a strong emphasis on the search for diagnostic characters which are to be used for delimiting taxa, rather than concentrating on the description of variation within taxa. Finally, since the phylogenetic species concept aims to identify all evolutionary taxonomic units, it results in a much more accurate assessment of inter- and intra-cladal diversity patterns than does the biological species concept.

Cracraft comments that one of the *dis*advantages of the phylogenetic species concept over and above the biological species concept might be that it recognises many more species level taxa, whilst one of Mayr's aims with the biological species concept was to introduce greater clarity and simplicity into biological classification. However, Cracraft adds that he does not agree that this is a disadvantage, since the phylogenetic species concept principally elevates what were rated as subspecies by the biological species concept to the level of species. He says:

> "If those 125 species represent our best estimate of the number of real evolutionary taxa in the family, then 40 additional species seems a small price to pay, especially when most, if not all, of those taxa will already

[172] There are other divergent views, such as that of van Valen, which takes speciation to be the process by which descendant populations are ecologically differentiated from their ancestors. (Cited in Kitcher, 1984.)

have been described and have valid names available."[173]

Hull transmits the pheneticists' feelings about phylogeny when he tells us that Sokal and Sneath admitted that phylogeny is "an all-explanatory principle" but said that it cannot be used in "classificatory procedures, since we mostly do not know (and in many cases cannot know) its true course."[174]

Pluralistic Species Concepts.

I now want to take a look at two accounts of pluralistic species concepts. These are Philip Kitcher's "pluralistic realism"[175] and John Dupré's "promiscuous realism".[176] Both recommend radically pluralistic accounts whereby the variety of existing species concepts is acknowledged and no one is given priority over and above the others. The argument is that pluralism is justified since the biological world comprises many different aspects and different species concepts therefore reflect these different aspects of biological reality.

Kitcher's Pluralistic Realism.

Kitcher claims that:

> "*Pluralistic* realism rests on the idea that our objective interests may be diverse, that we may be objectively correct in pursuing biological enquiries which demand different forms of explanation, so that the patterning of nature generated in different areas of biology may cross-classify the constituents of nature."[177]

He prefaces his account with a list of the various different species concepts and by adding that, "...in each case, the champions of the proposal contend that their species concept can serve the purposes of all biologists. In this I think that they err."[178] He then goes on to list some of the problems associated with the biological species concept – namely that is is impossible to evaluate whether two forms which have been extinct for some time were reproductively isolated or not and that the biological species concept is inapplicable to asexual organisms. Kitcher believes that such problem cases must be taken seriously and not simply dismissed as infrequent

[173] Cracraft, 1992, p.106.
[174] Sokal and Sneath, 1963, p.265 (quoted in Hull, 1988, Chapter 4).
[175] Kitcher, 1984.
[176] Dupré, 1981 and 1993.
[177] Kitcher, 1984, p.330.
[178] Kitcher, 1984, p.317.

anomalies. He points out that, "Although the biological species concept brings out an important pattern in the diversity of nature – the division of organisms into groups that are reproductively isolated from one another is theoretically significant – this is not the only important pattern of organismic diversity." and, "[the problem cases] point to distinctions among organisms which can be used to generate alternative legitimate conceptions of species."[179]

In order to explain how pluralism is possible, Kitcher makes the distinction between two different kinds of explanation in biology – *structural explanation*, which aims to explain the properties exhibited by organisms in terms of underlying mechanisms and structures and *historical explanation*, which aims to isolate the evolutionary forces which have produced the morphology, behaviour, ecology etc. of both extant and extinct organisms.[180] Kitcher emphasises that neither type of explanation is more important than the other, since each seeks to explain a different set of phenomena and so, "We should not confuse ourselves into thinking that one type of answer is appropriate to both types of questions or that one type of question is more "ultimate" than the other."[181] Consequently, he claims that these two kinds of biological explanation produce differing schemes for the classification of organisms – one scheme based on structural traits and the other based on historical (evolutionary) traits.

However, the pluralism runs deeper than this. It is not simply that two different types of classification are possible, but that within each of these schemes, there exist finer distinctions still. For instance, it may be possible to explain diversity amongst organisms by referring to structural traits at *different* levels – explanations involving genetics may be appropriate for some organisms, whilst explanations involving chromosomes may be more appropriate for other organisms.

Kitcher illustrates the aforementioned point by reference to the phylogenetic species concept, which groups organisms according to their recent common ancestry. In order to get this kind of classification off the ground, we need to employ a principle of phylogenetic division which will govern what changes are judged important enough to give birth to a new evolutionary unit. Kitcher suggests that there are three main candidates for this task – the production of branches which are reproductively

[179] Kitcher, 1984, p.319.
[180] This is based on Mayr's distinction between functional and evolutionary biology.
[181] Kitcher, 1984, p.321.

isolated, the attainment of ecological distinctness and the development of a new morphology. His argument is that these are "...[the] three important types of division among organisms, and each of these three types of division can rightly be used as the criterion for disrupting phylogenetic continuity or as a phenomenon of interest in its own right."[182] His point is that which principle of division the biologist adopts will (in part) be governed by his area of interest and by the nature of the organisms with which he is dealing. So, to take Kitcher's own example, a paleontologist who is reconstructing the phylogenies of classes of organisms may use the reproductive isolation of descendant branches to divide well-understood vertebrates into species, whilst he may use ecological or morphological discontinuities to divide asexual plants or marine invertebrates into species.

By way of conclusion Kitcher states that:

> "...we must recognise that there are many different contexts of investigation in which the concept of species is employed, and that the currently favoured set of species taxa has emerged through a history in which different groups of organisms have been classified by biologists working on different biological problems."[183]

Dupré's Promiscuous Realism.

Dupré's promiscuous realism is very close to Kitcher's pluralistic realism and Dupré states that he wishes to endorse Kitcher's view.[184] Promiscuous realism entails that:

> "There is no God-given, unique way to classify the innumerable and diverse products of the evolutionary process. There are many plausible and defensible ways of doing so, and the best way of doing so will depend on both the purposes of the classification and the peculiarities of the organisms in question..."[185]

One of Dupré's motivations for advocating species pluralism is that it is simply not clear that there is a *single* factor which can account for the cohesion of species across the board. He points out that the biological species concept cannot give a satisfactory account of the diversity of asexual organisms and that it therefore seems reasonable

[182] Kitcher, 1984, p.324.
[183] Kitcher, 1984, p.331.
[184] Dupré, 1993, p.51.
[185] Dupré, 1993, p.57.

to expect that there might be factors other than reproductive isolation which can account for diversity amongst *sexual* organisms and according to which divisions between species can be drawn. This would seem to provide a metaphysical ground for advocating species pluralism.

Dupré argues that Darwin's theory of evolution should force us to realise that nature does not possess a unique structure, hence our reasons for accepting a particular taxonomic scheme will be that it serves a particular purpose better than other possible schemes, not that that is *the* uniquely correct scheme of classification. He insists that he is ontologically (not methodologically) motivated to advocate pluralism and that, "It is that the complexity and variety of the biological world is such that only a pluralistic approach is likely to prove adequate for its investigation."[186]

The objection that such pluralism leads to a culture of 'anything goes' and to the abandoning of realism is anticipated by Dupré. He argues that nothing he says forces one to abandon realism and this is because realism with regard to biological kinds is quite separate from the belief that there is one single factor available to account for diversity amongst biological organisms. Simply because distinct and overlapping kinds are a possibility does not mean that realism will be lost and so, "...an organism might belong to both one kind defined by a genealogical taxonomy and another defined by an ecologically driven taxonomy."[187][188]

Finally, contrary to the weight of recent opinion, Dupré suggests a possible way in which morphological variation may produce biologically significant classifications.[189] This involves interpreting morphology in terms of properties which are highly significant in ecological situations. Recently favoured species concepts (the phylogenetic and biological species concepts) focus on the fundamental relevance of evolution in classifying biological organisms. Dupré insists, however, that there is more to biology than evolutionary theory, stating that, "It is just not the case that the parts of evolutionary theory that deal with the events and changes that species undergo is the sole theoretical context in which species figure."[190] He

[186] Dupré, 1993, p.53.
[187] Dupré, 1993, p.58.
[188] Kitcher makes similar attempts to defend pluralism against anti-realist criticisms. See his 1984, p.330.
[189] We saw that Kitcher also emphasised ways in which morphology is relevant for biological classification.
[190] Dupré, 1993, p.42.

considers the case of ecology where (he claims) the application of ecological theory to real-life ecological situations will be dependent upon a classification of the relevant organisms. He uses the example of a particular ecological system in which lynxes are classified as the predators and hares are classified as the prey, arguing that the relevant explananda here will be things like fluctuations of numbers in the species. He points out that an appreciation of ecology is in fact extremely important for a proper understanding of evolutionary development.

Dupré continues that it is acknowledged that the occupants of a specific ecological niche do not always coincide with the members of a specific genealogical line. In these kinds of case, it is argued, the ecologist will be justified in preferring a classificatory system which is not based on genealogy in order to reflect the clusterings with which he is concerned.

Dupré sums up:

> "Why assume that the play must be an evolutionary one? Will that not depend on...what time scale we are interested in? Ecological plays take months or years, while evolutionary ones take eons. The problem is that the actors in the short plays are on occasion different from the actors in the longer play that they jointly compose."[191]

In other words, there is more than one factor at work here and concentrating on either one will motivate differing but equally legitimate or real classifications.

Conclusions.

We have learnt several relevant things by studying modern biological species concepts and schemes of classification. Firstly, we have seen that there are number of different species concepts which are current in biology. The proponents of these competing schemes seek to establish their own scheme as supreme and all-purpose and to denigrate the others as being inadequate in some respect(s). It is important for my purposes to recognise that this debate is a live and very real one and not simply something to be relegated to the history of science with modern biologists having agreed unanimously on a single species concept. Secondly, we have seen that although it is convenient to align the various theories under the headings 'phenetics', 'biological species concept' and 'phylogenetic species concepts', still there is

[191] Dupré, 1993, p.43.

variation within these categories. This is particularly the case with the phylogenetic species concept, where more recent work diverges quite radically in some cases from the that of the original theorists. This just provides more grist for the mill in support of the argument that the debate surrounding species is very much alive and well. Thirdly, we have see that, in response to the variety of available species concepts, some scholars are advocating radical species pluralism, declaring that there is not one supreme classificatory scheme, but that different schemes serve different purposes and different subject matter better in some cases than others. This provides support for my own contentions, developed throughout this thesis, that the natural world is sufficiently complex that it does not come with a single preferred description, that the scientist is therefore forced to make a choice regarding which factors are relevant for matters of classification and that this can result in competing classificatory decisions, all of which are legitimate, since they each reflect different aspects of biological reality.

Having established that debates concerning taxonomic criteria and differing species concepts are anything but a relic of the past, I now want to look at three case studies taken from the history and sociology of science and to demonstrate that, when analysed, they corroborate a number of claims which I want to make concerning scientific classification.

"Controversy Over Classification: A Case Study From the History of Botany" – John Dean.

Dean's case study deals with plant taxonomy. He catalogues the controversies surrounding two very different methods of classification in the domain of botany; that of orthodox (Linnean) taxonomy and that of more modern experimental taxonomy or biosystematics. This provides a particularly interesting example, since the controversy between the two methods continues up to the present day. Neither has come to be considered absolutely right or wrong and so neither has been discarded. The natural world supports both methods and so it in fact makes no sense to talk about one as right and the other as wrong. Dean argues that this is extremely good evidence for the thesis that the classifications which are used in relation to the natural world are invented and not discovered and that those classifications will reflect certain professional vested interests and objectives.

Dean usefully begins by recounting the historical development of both methods of

classification. Orthodox taxonomy holds that classification must depend on morphological discontinuities between plants which are capable of being discerned through perception alone rather than through experiment. Adherents of this view believe that only by utilising such methods is it possible to arrive at (discover) a classification which reflects the inherent taxonomy of the natural world.

Orthodox taxonomy is essentially Linnean. That is, it is based on the essentialist precept that species are entities which exist objectively in the natural world and are individuated by possession of real essences or characteristics. The taxonomic aim then becomes one of discovering these real essences and so constructing a system of classification which is a mirror image of that existing in the world. The major consequence of this approach is that scant attention is paid to *intra*-specific variation. It is variation *between* species which is seen as of fundamental importance for matters of taxonomy.

The second facet of orthodox taxonomy is that it is eminently *practical*. It incorporates standardized descriptions, precise rules and enables accurate storage and rapid retrieval of information. It is designed to enable its users to classify plants with ease and in comfort, by examining the morphology of dead plant material in the herbarium. As a result, it is particularly easy to classify new species of plant and fit them into the confines of the established system.

Surprisingly, the orthodox system went unchallenged even with the general acceptance of Darwin's theory of evolution in the second half of the nineteenth century. Darwin argued that species develop gradually and that change from one species to another is too gradual to be able to delimit them. If this is so, the classical view of species as fixed and unchanging becomes untenable. However, this did not affect the methods used in traditional taxonomy. It was accepted that organisms were related by descent but still the taxonomists' aims remained the same – to arrange this diversity into some structured order on the basis of morphology alone. Even if species were fictions rather than realities, it was seen as practically necessary to use them as the basis of division.

Criticism of Linnean taxonomy first arose with the development of new theories and practices in fields outside taxonomy. Most notably, following the rise of knowledge about and interest in genetics (the study of heredity and variation in organisms) in the years following 1900, the new fields of (cyto)genetics and cytology (the study of

cells) came into being. These endowed taxonomy with a new set of techniques, such as transplanting and crossability experiments and detailed chromosome analysis which paved the way for an alternative taxonomy. In particular, the adherents of these new practices found that Linnean taxonomy was inadequate regarding description of intraspecific variation. The new focus of study became variation rather than species identification.

The new biosystematists argued that in order to produce a correct classification of the plant kingdom, it is necessary to adopt experimental methods and to make use of categories with an experimental base. Probably the most important experimental method has been the study of crossability, which investigates the limits of gene-exchange between populations in nature and in the laboratory.

Despite the enormously different aims and methods of the two systems of classification, they have continued to exist side-by-side; there has been no abandonment of either the one or the other. Orthodox taxonomists tend to defend their methods on practical or utilitarian grounds. They agree that the evolutionary aspects of a plant's history are very important, but deny that this makes a good general basis for classification in terms of practicality, since the process is a slow one, requiring much research and experiment.

Since both systems continue to be used, competing classifications have grown up. Often, use of either method will result in roughly the same classification. However, there is conflict, especially at the species level, where the two methods make use of different definitions. In traditional terms, species are populations of morphologically similar entities which are separated from other species by morphological distinctions. In biosystematics, a species is a population of individuals which interbreed and are capable of gene exchange. Most frequently, conflicting classifications occur concerning plants with asexual or partly asexual breeding cycles.

A Case Study of a Classification Controversy.

Dean goes on to recount two specific instances of controversies over classification which are created by the two conflicting systems. He examines the plant genus Gilia and the problems concerning two species-complexes within this genus – *Gilia inconspicua* and *Gilia tenuiflora*.

The Gilia inconspicua complex comprises no less than five interrelated sibling

species. It can be shown that each of these, although fertile in themselves, are all highly intersterile. From the point of view of the biosystematist, then, all five form separate species. However, morphologically, all five form a single group, since it is only by dint of microscopic examination of chromosomes that these five can be distinguished. Thus, where the orthodox taxonomist would claim only one species, the biosystematist would claim five separate species.

Likewise, the Gilia tenuiflora-latiflora complex comprises at least four different elements which are quite morphologically distinct, yet are capable of gene exchange. Thus, the orthodox taxonomist would claim at least four separate species, whilst the biosystematist would claim only one.[192]

Dean concludes, "Nor should it be imagined that examples like the one given here are an exceptional occurrence and hence of little importance to most systematics...such groups are the rule rather than the exception in higher plant taxonomy."[193]

Evaluation of Dean's Case Study.

I suggest that Dean's case study presents a very interesting piece of evidence for the thesis that there is more than one possible division of the natural world, that essentialism is misguided and that any taxonomy involves input from both the world and from the human beings who inhabit and cognise that world. Dean's work is particularly valuable, since it provides us with an example of a current and unresolved conflict concerning competing taxonomies within the realm of science. It is unusual to come across such a conflict which has not yet come down on either one side or the other.

The Inherent Messiness of Science.

Firstly, Dean's study shows that there *can* be controversy within science concerning categorisation. This is something which I believe many people fail to realise. They assume that if they asked a botanist how a particular plant should be classified, he would be able to give a conclusive answer, together with the reasons for his answer. It would further be assumed that he would be speaking for all botanists. I suggest that it is unlikely that people have a conception of there being different ways in which

[192] See Dean, 1979, for a technical biological explanation of the differences between the sibling species.
[193] Dean, 1979, p.224.

one might classify a plant or that there are different scientific reasons which might be proffered for why that plant is to be classified in one particular way. This is because this kind of scenario is quite at odds with the picture of science which is generally presented to the layman – that of a neat and tidy affair with only one scientifically correct answer, given the way the world is.

Human Input for Categorisation: Aims, Theories and Commitments.

Another very interesting facet of Dean's example is that we are not talking about a controversy as to how a particular plant should be categorised within one taxonomic system, rather, we are talking about two entirely different systems of classification and controversies arising from the use of the different systems. The point to be appreciated here is that the adherents to each system have different interests; different things are important to them professionally. What this means is that which system a botanist adopts will be dictated by his prior interests and concerns. Adherents to orthodox taxonomy are aiming to introduce order into the natural world and to produce a highly practical taxonomy. They are interested in drawing sharp boundaries between species and making use of precise definitions and rules. They believe that the way to produce a correct classification of the natural world is to make use of easily perceptible morphological discontinuities between groups of individuals. This enables them to remain working in the herbarium with dead plant materials, on the basis of which they can make quick and easy classifications. It is all these aims which underlie the choice of orthodox over and above experimental taxonomy by certain individuals. The very system itself was in fact adopted with those particular aims or ideals in mind.

On the other hand, the biosystematists (experimentalists) see morphological discontinuities as subservient to certain experimental techniques which highlight differences in interbreeding and gene exchange. They are far more interested in intra-specific variation that in specific discontinuity. They do not consider quickness and ease to be important or relevant factors in the taxonomist's classification. Their aims and interests lead them to assert that the only way in which to produce a correct or satisfactory classification of the plant kingdom is to use experimental techniques – to base classification on a very different set of precepts from those of the orthodox taxonomist. Again, this very system of classification grew up and was developed with certain very definite aims and interests in mind.

The fact that these two quite different systems of classification exist side by side is evidence of the human input which I argue is always involved in matters of categorisation. It appears that there is an element of choice at work here; there are different ways according to which we might divide up the natural world. Dean's study is very clear on this point. Here we have a real and current situation where two different systems exist which are based on different sets of aims and interests. Whether you are orthodox or experimentalist will depend on a prior decision as to what factors are important in order to produce an adequate classification of the plant kingdom. Botanists are, quite clearly, bringing something of themselves and of their beliefs to their work. These prior commitments then serve to highlight certain regularities or similarities in the world, which then form the basis of a categorisation system. The biosystematist's interest lies in certain experimental techniques. These techniques serve to highlight or make salient facts about interbreeding and gene exchange. These facts or regularities then act as the basis upon which divisions between plants are drawn. The orthodox botanist, on the other hand, is interested in making quick and practical classifications on the basis of the examination of dead plant material. He is able to do this by concentrating on morphological discontinuity – this is what is salient for him. Facts about differences in morphology between plants are thus used as the basis for dividing those plants into classes whose members are judged to be similar.

The interesting thing is that although the aims of the two systems and some of the classifications which they produce are competing, it makes no sense to label one system as correct and the other as incorrect. Dean himself tells us:

> "If classification is a process of discovery, then a direct appeal to nature should be sufficient to evaluate the best taxonomy. Is this in fact the case? Evidently, no. In this case, both taxonomies are conventions designed to emphasise different aspects of the real world. Consider again Gilia inconspicua: morphologically, we do indeed find only a single species: experimentally, evidence can be accumulated by which five different species may be discerned. Both taxonomies are built upon perceptible, systematizable, stable distinctions between individual plants. In this sense the natural order sustains both taxonomies; neither can be said to involve a distortion of the real facts; neither can be said to be erroneous. Nature does not in itself allow such an evaluation to be

made."[194]

The point, then, is that the natural world is able to sustain both systems of classification. In other words, it is divisible in more than one way. Apparently, there is not one ultimately correct taxonomy of the world, but there may well be several, each of which may be justified to the same degree as the next.

Categorisation as More than the Way the World Is.

How is it possible that the natural world can support more than one system of classification? As outlined above, classical and experimental taxonomists both have very different aims and purposes in mind and as a result of these differences, produce differing classifications of the plant kingdom. These aims and interests dictate that certain patterns which exist in the environment will be salient or relevant for matters of classification. For the biosystematist, facts about interbreeding and gene exchange become salient; for the orthodox taxonomist, facts about morphological discontinuity become salient. It is important to grasp that these salient facts actually reflect *different facets of* reality. Plants do, *in reality*, exhibit morphological discontinuities. They do, *in reality*, also exhibit certain patterns concerning inter-breeding and gene-exchange. The natural world is able to support both methods of classification simply because it is an inherently complex system with more than one pattern of similarity or regularity running through it which may successfully act as the basis of division.

Each group of taxonomists latches on to one particular facet of reality, making this into the basis of their classification system. This is not, however, a random or arbitrary process, since both morphological discontinuity and interbreeding and gene exchange are objectively, naturally existing phenomena which form regularities in terms of which botanists are able to interpret the natural world. As Dean says, they are both "perceptible, systematizable, stable distinctions".[195]

What this goes towards illustrating is that the classifications which scientists make are not simple reflections of reality in itself. Rather, classification represents an admixture of aspects of the natural world plus prior theories, aims and commitments which are brought to the world by the scientist. Both the classical taxonomist and the experimentalist approach their work with a prior schema in mind and it is their prior aims which determine which patterns in the natural world they will pick out as being

[194] Dean, 1979, p.226.
[195] Dean, 1979, p.226.

relevant for matters of categorisation. Categorisation is therefore not a foregone conclusion dictated by the way the world is. Rather it involves evaluations and choices, reflecting prior aims and commitments, which result in the selection of certain natural regularities over and above others.

Dean himself states that his research shows that, "classification is best seen as a process of invention rather than discovery".[196] Although I sympathise with his general position, I think his use of the word "invention" is open to misinterpretation. It tends to be seen as carrying negative connotations of arbitrariness and non-objectivity. However, as I have taken care to point out, the fact that two competing classification systems which frequently yield different results can exist side by side is *not* to say that they are not based on objective regularities or patterns. Indeed both systems represent entirely real and objective regularities in nature, it is just that reality is sufficiently complex and rich to be capable of supporting systems which are based on *different* instances of regularity. I am sure Dean himself has no illusions that these systems are in any way arbitrary, since he speaks of the two systems portraying "different aspects of *reality*"[197] and of the fact that, "Both taxonomies are built upon perceptible, systematizable, stable distinctions between individual plants."[198]

The Problem with Essentialism.

On the Putnamian-Kripkean conception of natural kinds and essentialism which I discussed in Chapter Two, the job of science is pictured as providing us with a mirror image of the natural kinds in the world, together with the essences which individuate those kinds. There is no suggestion that nature may allow alternative carvings of the world or that scientists may come up with conflicting, yet equally valid taxonomies or that categorisation may involve any kind of synthesis of subjective and objective elements.

Dean's case study proves the essentialist wrong on several counts. Firstly, it shows that the scientist is not simply engaged in the disinterested pursuit of reflecting the world's inherent natural kinds, together with the basis for their distinction (essence). Rather, the scientist brings certain prior commitments to his work, which shape the way he thinks and operates. Secondly, it illustrates that there is no such thing as one single ultimately correct classification which reflects the world's inherent division.

[196] Dean, 1979, p.226.
[197] Dean, 1979, p.225. My italics.
[198] Dean, 1979, p.226.

Rather, reality is sufficiently rich and complex to support competing systems of classification, each of which can be judged to be equally objective, correct or real. Thirdly, it shows that so-called natural kinds and their essences are by no means a simple or absolute matter. Rather, which groupings one perceives in the world and the essences which one assigns to those groupings will be dictated by ones's prior commitments and theories, one's prior evaluations of what is salient and important for categorisation. For the experimentalist, the essence of a particular group will comprise facts about gene exchange between members of the group. For the orthodox taxonomist, the essence of a particular group will comprise possession of a common morphological structure.

The Blurring of the Line between Metaphysics and Epistemology.

This kind of evaluation of Dean's case study also illustrates my contention that, as regards categorisation, there is a blurring of metaphysics and epistemology and that it is, in fact, impossible to draw any kind of sharp dividing line between the two. This is because, as we have seen, the natural world is able to sustain alternative and competing systems of classification. Although each system reflects a certain aspect of *reality*, the decision as to which system a scientist uses will be prompted by prior commitments and aims. These commitments and aims will cause the scientist to pick out certain patterns or regularities occurring in nature as the salient ones for producing a successful (and, from his point of view, correct) taxonomy of the natural world. Thus the decision as to *which* regularities to base categorisations upon cannot be seen to be a matter of metaphysics. Rather, this is a matter of epistemology; it is a matter of choosing certain things over others on the basis of prior commitments to aims and theories which dictate how the scientist expects that the world should be divided. At this level, epistemology has a profound affect upon metaphysics. Dean's study therefore shows us that scientific classification must be viewed as a synthesis of both metaphysics and epistemology.

Support for an Explanation-Based Model of Categorisation.

It seems that Dean's case study is well-placed to provide good evidence for my claim that the psychological explanation-based view of categorisation is a suitable model for classification by scientific experts as well as by laypeople.

In fact, Dean's study fits rather neatly with the requirements of the explanation-based view. According to his evidence, how taxonomists categorise the plant kingdom will

depend on whether they are using an experimental or a traditional system, which in turn depends upon what they take the fundamental goal of taxonomy to be. Following the explanation-based view, it is not adequate to say that the things which we class together are similar *per se* and that this is what grounds our classification. Rather, it is felt that similarity can only be said to make sense if it is grounded in some sort of theory or explanation, the reason being that objects can be said to be (dis)similar in all manner of ways. A theory or explanation is needed in order to explain *why* certain properties are taken as being salient for judging two or more individuals to be similar. Likewise for classification; to say we group things together because they are similar will not do. Rather we need to say *why* or *in what context* (under what theory or explanation) these objects count as similar and therefore are judged to belong to the same class or category.

In the case of orthodox and experimental taxonomy, we have two different classification systems which are capable of producing conflicting groupings. In order to make sense of the conflict, we must invoke the appropriate sets of theories, aims and explanations. The plants are not simply similar *per se*, rather certain similarities can be seen to be salient in terms of a particular theory and set of commitments. For the traditional taxonomist, morphological discontinuity is seen as the relevant factor, the aim being to produce a quick and easy taxonomy based on easily perceptible forms. For the experimentalist, facts about interbreeding and gene exchange are seen as relevant or salient, the aim being to concentrate on matters of evolution and speciation. The different theories, aims or explanations underlie the conflicting classifications which are produced. It is not appropriate to make a judgement regarding the truth or falsity of either taxonomic system since both systems are supported by reality. Each invokes an alternative pattern of similarities which may be discerned in nature.

Dean's account supports a second point made by the explanation-based theory – that concepts are relational entities which gain much of their meaning from their association with other concepts and from the role which they play in the context of the wider world. They are not simply isolated individuals whose meanings can be explained in terms of lists of defining conditions. We learn from Dean that in the domain of botanical classification, judgements of similarity and so category membership are made on the basis of certain prior commitments (explanations or theories). In this case, to fully appreciate what is going on when a botanist classifies

a plant, it is necessary to take into account these prior commitments in order to see why certain patterns of similarity are judged to be more fundamental or salient than others and so why a particular plant is included within a particular class. The underlying theory/explanation plays a vital role in understanding decisions regarding category membership since it provides the context in which these decisions make sense and come to life. For example, for the traditional taxonomist, a particular species will be delineated on the basis of morphological similarity and dissimilarity, whilst for the experimentalist, it will be delineated on the basis of interbreeding and gene exchange. Until we start talking about why different botanists pick out these differing factors as relevant, it is impossible to have a proper grasp of the different species concepts, since we do not have a full understanding of the background contributing to their meaning. Therefore, to concentrate upon the criteria for class membership which have already been chosen and to ignore the prior theories and commitments which have driven the choice of those similarities and those classifications is to ignore half the story. In order to give a full account of how classification works, we must not ignore this element of the issue. If we do, we will be liable to fall into error, to suppose essentialism is true and to assume that there must be a unique division of the natural world.

This ties in with one final point made by the explanation-based theory – that concepts and categories comprise much more than lists of attributes. They must also involve explanations of why those particular attributes are relevant, how the attributes relate to one another and how category members relate to one another. As explained above, Dean's study shows that botanical classifications involve much more than simple lists of attributes. He shows that in order to make sense of these classifications, we must take into account the factors which explain *why* particular attributes are chosen as being more salient or relevant than others.

"Designing the Dinosaur: Richard Owen's Response to Robert Edmund Grant" – Adrian J. Desmond.

This case study is slightly different from the last. Desmond's study deals with the historical 'creation' of a category (the dinosaur) by one Richard Owen. He aims to explain the subjective factors surrounding the introduction of the category and he stresses that its introduction was not just a case of discovery on the part of Owen.

Desmond shows that Owen's dinosaurs were quite different from the animals which

those working immediately prior to him had conjectured. He argues that in order to understand why Owen envisaged these animals so differently and why those researchers going before him failed to 'see' what he 'saw', we must inform ourselves of Owen's personal esoteric motives. Desmond stresses that Owen, "did not simply recognize dinosaurs, giving taxonomic expression to the 'objective' fact; more properly, he designed them – invented them, in a sense..."[199]

Prior to Owen, there was a healthy paradigm in place which saw Owen's dinosaurs as having been enormous reptiles, called "Fossil Lizards". This paradigm was based on the evidence of fossil jaws and teeth which indicated similarities to extant lizards. However, in his "Report on British Fossil Reptiles" made to the British Association in 1841, Owen elevated these creatures to the status of *Dinosauria*. In other words, he placed them in their own unique taxonomic category. He made reference to anatomical peculiarities of the sacrum, ribs and extremities and to their enormous size, which helped to distinguish them from living lizards and from Mesozoic marine saurians (lizards). From these references to anatomy, Owen went on to make bold conjectures about the vital ecological significance of his dinosaurs. Furthermore, although he referred to these creatures as cold-blooded, he insisted that they were akin to pachydermal mammals (large thick-skinned mammals, such as the elephant and rhinoceros), claiming that since they had the same thoracic structure as crocodiles, they must have possessed a four-chambered heart and that since they were so well adapted to life on earth, they must have had a circulatory system close to that currently possessed by warm-blooded vertebrates. Yet these animals for which Owen made so many hypotheses were only scantily known on the basis of three species.

Desmond goes on to tell us that estimates of dinosaur size had been made using the lizard paradigm. This was done by comparing the fossil bones with those of contemporary lizards and then scaling up accordingly. However, since femurs were often larger than an elephant's thigh bone, the estimated sizes were enormous.

Owen was doubtful about the above method of measurement and so constructed his own, whereby he measured individual fossil vertebrae and estimated their total number by using existing crocodiles and lizards as prototypes. This yielded slightly more conservative size estimates. However, in the process of so-doing, he managed to effect a gross morphological transformation of these animals. He replaced the

[199] Desmond, 1979, p.224.

paradigm of small lizard-type legs with mammalian-type legs and claimed that due to their massive weight, his dinosaurs would have stood upright on all fours, rather than sprawling close to the ground. Again, this was a puzzling conjecture, given that Owen had so little evidence to support his position.

Why was Owen so keen to claim that his dinosaurs were in fact much more akin to mammals than reptiles, in opposition to his predecessors and on the basis of scant evidence? Desmond suggests that it was due to Owen's personal antipathy towards Lamarckism in general (also known as transmutation in this context) together with his intense personal dislike of one eminent follower of Lamarck, Robert Edmond Grant.[200]

Lamarck had imagined a huge sequence of life forms extending from the simplest to the most complex, with human beings at the pinnacle of perfection. He believed that animal organs were possessed of some kind of internal energy or excitation, which impelled these organs to become more complex, with the result that life forms gradually took their place on successively higher levels. Two 'laws' were said to govern this system; organs were supposed to be improved with repeated use and weakened by disuse and the consequent changes were said to be preserved in reproduction. This system can be seen as one product of the notion of a metaphysical chain of being. Grant's aim was to provide an historical construction of Lamarckism. He believed that evidence of Lamarck's system would be imprinted on the fossil remains; that the earliest fossils would be of forms at their most primitive, with later fossils showing a gradual rise in the complexity of these forms.

Owen actually tested Grant's hypothesis, but found no support for it. The New Red Sandstone rocks should have been home to the most primitive and ancient amphibians, yet the fossil forms found there were clearly more advanced than currently existing salamanders, frogs and apodans. It seems that, contrary to Grant's Lamarckian hypothesis, these forms first made their appearance in their highest rather than their lowest developmental form. Furthermore, Owen suggested that the fossil evidence showed that lizards were actually more ancient than the amphibians which were supposed to be their predecessors, according to the Lamarckian scale.

[200] Desmond informs us that Owen was incensed by Grant's attempt to construct an historical Lamarckism and that they came into conflict publicly in the 1830s. Owen wielded much power in both science and society and in the mid 1830s he prevented Grant's appointment as comparative anatomist at the Zoological Society of London. When dealing with the Lamarckian problem, Owen quoted from some of Grant's lectures, considering him to be the latest upholder of transmutation in relation to paleontology.

Desmond goes on to argue that Owen used the example of the dinosaurs to further the case against transmutation theories. According to Owen, reptiles at their peak (as dinosaurs) approached the complexity of the mammal form both morphologically and physiologically. Yet this occurred way back in the Mesozoic era. If this was so, then it appears that the reptiles later suffered a degeneration, resulting in the contemporary large numbers of small lizards. Owen thus effectively refuted the transmutation argument for a continual increase in fossil complexity through time.

With his dinosaur evidence, Owen began to argue for degeneration rather than for transmutation. This fitted well with opinions of the time. Transmutation theories were seen as a threat to Christianity, since they seemed to suggest materialism, with each life form having *within itself* the power of self-improvement. In addition to this, evidence such as that produced by Owen seemed to suggest that although species had a highest form, they might proceed to degenerate from that rank. In response to this, a compromise position called "Discontinuous Progression" was eventually reached. This was based on the observation that some early members of some classes exhibited high complexity. It was argued that life did advance, but step-by-step rather than continuously. Each class was said to appear at a discontinuity and, once created, was said to stay at that level or even to decline. Since there was no continuous link between classes, it was claimed that God must have intervened on each occasion to create each new class. Thus, succession of classes, degeneration and the the omnipotence of God were all preserved within this new theory. However, transmutation theories themselves were rejected out of hand; they failed to account for degeneration and threatened the supremacy of the Christian tradition.

Desmond concludes that Owen had the means and the motive for creating the dinosaur by reinterpreting the fossil evidence and overhauling the fossil lizards. He had the means because degeneration was becoming the standard response to transmutation. His motive was his great dislike of Lamarckism, plus his personal hatred of Owen. All these factors combined to prompt Owen to 'create' the dinosaur and together explain why Owen rather than those researchers going before him exploited this possibility. As Desmond says, "...[it is assumed] that science is a search for transcendental truths; timeless truths which, being everpresent, passively await the man perspicacious enough to recognize them. It is more profitable, however to view science as a culture-bound, inherently creative activity. Problems only arise under certain conditions; it is these problems peculiar to the age which call

forth novel responses, and it is this relationship which we call science."[201]

Evaluation of Desmond's Case Study.

Desmond's study is somewhat different from that of Dean. Dean's study showed us how and why there are currently two different systems of classification used in botany and illustrated that these systems sometimes produce competing classifications. Desmond's study is not concerned with entire taxonomies or systems of classification, rather with the historical emergence of a new category (*Dinosauria*) in paleontology and the nature of the animals said to fall into that category. He is concerned to show us that the emergence of such categories is not merely a matter of discovery of objective facts which are then catalogued accordingly, rather it is a much more complex process involving certain socio-historical situations together with the motives and means available to individual scientists. Desmond makes it clear that he intends this kind of analysis to be applied to the whole of science and not just to this specific case.

The Inherent Messiness of Science.

Desmond's case study is interesting in that it paints a good picture of the inherently 'messy' nature of science. It provides support for the notion that the public image of science as objective, clear-cut and prejudice-free is far from the true situation. In reality, science involves prejudice, prior commitment and a degree of subjectivity. As we have seen, the very reasons for Owen's redesign of the dinosaur were based on personal preferences at least as much as any notion he might have had of searching for 'the truth'. He was driven primarily by dislike of Lamarckian principles and even more so by hatred for his rival Grant and his consequent desire to ruin this man's career. Had it not been for these personal feelings, it seems unlikely that Owen would have gone to such lengths to redesign the dinosaur. Furthermore, having decided to overthrow the lizard blueprint, he went beyond the fossil evidence available at the time, making virtually unsupported hypotheses regarding these animals in order to support his reconstruction of them. On the basis of some similarities between dinosaur bones and those of certain mammals, Owen made a huge leap, speculating about dinosaur ecology, physiology and soft anatomy and completely reconstructing their morphology. These were bold statements based on highly inadequate evidence. Yet, his arguments were accepted because the time and public opinion were right. Owen's speculations supported degeneration and militated against transmutation,

[201] Desmond, 1979, p.234.

thereby fitting neatly into the contemporary climate. As Desmond is keen to point out in his paper, science is in fact not characterised by the search for "transcendental...timeless truths", but is an "inherently creative activity".[202]

Of course, one of the ways in which science can be shown not to be characterised by the search for "transcendental timeless truths" is in the domain of classification. As Desmond's study shows, the natural world does not provide us with such truths to be used in classification. Hence, scientists are able to produce competing, yet equally objective or real groupings of entities in the natural world.

Categorisation as More than the Way the World Is.

We can use Desmond's case study to illustrate my contention that the way scientists categorise objects in the world is not purely a matter of how the world is in itself, of reflecting the world's inherent taxonomy. As we have seen, Owen's predecessors classified his dinosaurs as giant lizards on the basis of fossil evidence, particularly jaws and teeth, which seemed to resemble those of currently existing lizards, although much larger in size. Owen, however, rejected this paradigm and claimed that these animals much more closely resembled the pachydermal mammals in terms of their size and certain anatomical features. Desmond actually tells us that those researchers working immediately prior to Owen, "...all noted isolated anatomical similarities between "Fossil Lizard" and mammal bones, [but] they persisted in using the lizard blueprint for reconstruction."[203]

The point to be borne in mind here is that both of these classifications can be seen to be supported by reality. That is, in some respects these animals can be interpreted as being similar to currently existing lizards, and so can successfully be classified as giant lizards. Fossil evidence supports this kind of similarity and classification. On the other hand, these animals can in other respects be said to be similar to currently existing pachydermal mammals and dissimilar to lizards, which merits them being placed in their own class – dinosaurs. Fossil evidence also supports this kind of similarity and classification.

This leaves us in a position somewhat akin to the one that Dean's study left us in. It appears that our classifications are actually underdetermined by the way the world is. How to classify natural entities appears to be an open question. It seems that both the

[202] Desmond, 1979, p.234.
[203] Desmond, 1979, p.229.

giant lizard and the (Owen) dinosaur interpretation have some support in reality.

In other words, it is apparent, once again, that science must be seen as involving much more than the production of a correct taxonomy of the natural world which merely reflects the unique category structure which is inherent in the world. Rather, when scientists categorise entities in the natural world, they pick up on certain similarities which exist in the world and make these salient for matters of classification. The natural world is sufficiently complex that there is more than one set of similarities on which categorisations can be based. We can see this principle in action in Desmond's case study. Different similarities were taken as salient by Owen and his predecessors and used to make different classifications; one as 'lizard', along with all current and Mesozoic lizards, the other as a separate taxonomic group labelled 'dinosaur' (meaning an animal which resembles existing mammals in certain respects which serve to distinguish it taxonomically from both existing lizards and from Mesozoic marine lizards, such as plesiosaurs and icthyosaurs). Both classifications are successful in the sense that they are both grounded in reality; both are based on similarities which can be found in the natural world. It is not a case of deciding which of the classifications is 'correct' – each is correct in the sense of being equally real and an equally fair representation of the natural world, it is just that each represents a *different aspect* of that world.[204]

Human Input for Categorisation: Aims, Theories and Commitments.

The immediate reason why Owen and his predecessors were able to come up with competing classifications of the Mesozoic animals was that they saw different sets of similarities as important or salient and the reason for this is that they had different

[204] It is also worth pointing out that the different ways in which the Mesozoic creatures were classified by Owen and by his predecessors cannot be explained away by saying that there was not sufficient evidence for a decision to be made as to which was the correct classification, that if we had been around in the Mesozoic era and the evidence was before our eyes, so to speak, we would have been able to say how these animals should be correctly classified. These days we have considerably more fossil evidence. Dinosaurs are currently considered to have been members of a (now extinct) race of saurian (resembling lizards) reptiles. The fossil remains point to organisms which in some respects resembled birds and in other respects resembled mammals. In other words, the dinosaurs were similar to lizards, mammals and birds. Each set of similarities exploits a different aspect (or aspects) of these animals. Each similarity is as real as the next. It is just that the reptilian features of the dinosaurs have been taken as primary for matters of classification and so they have been grouped with the reptiles. Taxonomists might, however, have taken one of the other sets of characteristics as the most salient and so have classed the dinosaurs differently. Presumably, if we had been alive at the time of the dinosaurs, we would have been able to see that all these similarities were real, that each one could be supported by the world or by the facts. The situation is still the same, whether dealing with fossil evidence or with extant species. In either case, the natural world can support differing classifications.

aims and motivations which made those particular similarities salient for them. Desmond does not spell out the factors which motivated Owen's predecessors to classify these creatures on the basis of their similarity to lizards, since his paper concerns the reasons behind *Owen's* classification. All we are told is that Cuvier effectively set the lizard paradigm by judging a set of jaws found in Holland in 1795 to have belonged to a huge marine monitor lizard. It is possible to speculate about what the motivation for Owen's predecessors was. Perhaps Cuvier made his judgement on the basis of pure resemblance. If the jaws really did resemble those of existing lizards and the jaws were the only evidence he had, then it would be reasonable to make such a judgement. Once the tradition had been set, however, it continued. This would seem to be a case of entrenchment – the paradigm had been accepted, it was workable and could be used as the basis for making further judgements, such as size estimates of these animals. The second reason which can be put forward for why Owen's predecessors classified these animals as lizards is that they (perhaps even unthinkingly) had no quarrel with transmutation theories. In other words, they had no problem with Lamarck's idea of reality as a moving escalator of living forms. Unlike Owen, they had no reason to show that these animals were far more advanced than currently existing lizards, thereby illustrating that there must be something fundamentally wrong with transmutation. As Desmond remarks at the end of his paper, "One really cannot talk of "missed opportunities" on [Owen's predecessors'] part; the opportunity was never present, because the conditions that determined the necessary motivation were lacking."[205]

On the other hand, as we have already seen, Owen had plenty of motivation for altering the blueprint for these creatures, thereby creating an exclusive taxonomic category for them. He disagreed strongly with transmutation theories. By claiming that his dinosaurs in some respect approached mammals on the Lamarckian scale of things, he was able to show, by the lights of Lamarck's own principles, that a case of degeneration had actually taken place, thus throwing doubt upon the truth of transmutation theories. His case was easily accepted. General opinion had been turning against transmutation for some time because the theory was construed as a threat to the primacy of a divine creator and degeneration was already becoming the standard answer to transmutation. Owen was further encouraged in this project by his antipathy towards Grant, who had given a Lamarckian interpretation of the fossil record. In one fell swoop Owen, by creating the new category 'dinosaur', was able to

[205] Desmond, 1979, p.234.

accomplish two of his major aims – to help overthrow transmutation and to help discredit his enemy Grant. Owen's hatred of the Lamarckian system and in particular of one of its major proponents comprised the motivations behind his dinosaur classification. As Desmond comments, "...given Owen's vendetta against the University College professor, and his abhorrence of Grant's historical Lamarckism, together with this accepted manner of disposing of it, he possessed both the means and the motivation for overhauling "Fossil Lizard" morphology to add one more nail to the transmutationist coffin."[206]

As has already been discussed, the similarities which Owen urged between his dinosaurs and pachydermal mammals were supported to a certain degree by fossil evidence and by comparable size. The interesting thing is that Owen's predecessors were aware of and acknowledged the similarities which Owen was basing his speculations upon. Desmond tells us that, "Whereas Cuvier, Buckland, Mantell, and von Meyer all noted isolated anatomical similarities between "Fossil Lizard" and mammal bones, they persisted in using the lizard blueprint for reconstruction...Unlike his predecessors, Owen *used* this mammal quality, which he amplified and extended (speculating on dinosaur ecology, soft anatomy and physiology) in his polemic against the Lamarckians."[207] This quotation serves to reinforce two points. Firstly, that the similarity of dinosaurs to mammals was a reality and that even those who chose not to exploit it were aware of it. Secondly, it shows that which similarities are exploited by which scientists will depend on the underlying aims, motivations or theories of those scientists. Cuvier *et al* had no reason to exploit the mammalian similarity. However, Owen had a very pressing reason to exploit it and was able to do so successfully in order to accomplish his aims.

We can now see that Desmond's case study lends support to my claim that categorisation involves a degree of human input, decision and choice. We have seen that the world is able to support the competing classifications of these Mesozoic animals, since each classification gives credence to a different aspect or facet of reality. *Which* aspect each scientist chose to acknowledge was dependent upon his aims and beliefs. Whilst Owen was motivated by his beliefs about Lamarckism, this question simply was not live for or considered relevant by his predecessors. It is these aims and beliefs which force a choice or decision between similarities and so

[206] Desmond, 1979, p.233.
[207] Desmond, 1979, p.229.

comprise the element of human input for categorisation. If there were no such human input, then classification could not get off the ground. This is because the natural world has too many patterns of similarities running through it to support a unique categorisation of the entities within it, quite independent of human categorising agents. Once you have reasons for preferring one set of real similarities over and above another, then you have a basis on which to build your classification, but without this kind of basis, classification is impossible.

The Blurring of the Line between Metaphysics and Epistemology.

As with Dean's study, we seem to have evidence showing that in matters of classification, it is impossible to draw a sharp distinction between metaphysics and epistemology. We have learnt that Owen's dinosaurs can be classified in alternative ways since the natural world is sufficiently rich and complex and has enough different patterns of similarities running through it that it is able to support both classifications. If this is so, it cannot be the case that there is one ultimately correct taxonomy of the world, the job of science cannot be to reproduce that taxonomy and classification will not purely be a matter of metaphysics, since the world in itself fails to give precedence to any one pattern of similarities. Desmond's case study shows that the two different classifications were prompted by quite different sets of motives and aims. Only by approaching his work with some kind of prior goal or commitment in mind does the scientist have a tool with which he is able to single out one thread of similarities as the most salient or relevant for matters of classification. In other words, epistemology (in the form of some kind of human input) is involved in categorisation as much as metaphysics. In order to understand a classification you need to be aware of the scientist's motives and aims just as much as the ways in which the world is. Classification is thus a synthesis of metaphysics and epistemology; a synthesis of some of the many similarities in the world together with the reasons for choosing those similarities over and above others. Any attempt to draw a sharp line between metaphysics and epistemology and to claim that categorisation is purely a matter of metaphysics is reductionist and misguided. It risks completely missing the real complexities that classification involves and also risks assuming that one of the ways in which the world may be carved up qualifies as the correct way, whilst the others are incorrect or even do not exist as possibilities.

The Problem with Essentialism.

If I am right about not being able to draw a sharp distinction between epistemology

and metaphysics, then the kind of essentialism attached to natural kinds espoused by philosophers such as Putnam and Kripke must be mistaken. Desmond's case study shows that rather than science correctly reflecting the unique taxonomy of the world, it often actually produces competing classifications. It would be incorrect to use the terminology of right and wrong here, since both competitors exploit similarities existing in the natural world and the world fails to give precedence to either one of the possible groupings. Desmond's study shows that the notion of one uniquely correct taxonomy of the natural world is an *a priori* myth. Once one actually takes the trouble to examine the mechanism of scientific classification in detail, one begins to realise that the picture painted by the essentialist of the role and procedures of science is an inappropriate and incorrect one.

Support for an Explanation-Based Model of Categorisation.
Desmond's case study provides further evidence for my claim that the explanation-based model applies as much to scientific categorisations as to those made by laymen.

The study makes clear that categorisation cannot be explained in terms of similarity alone, since the natural world is far too complex to be pertinent in one way only. We must therefore ask *why* certain similarities are judged to be more important than others and so why certain similarities and not others are used as the basis for categorisation. As we have seen, Owen and his predecessors took different sets of similarities to be pertinent in classifying the Mesozoic creatures, which resulted in competing classifications. The reason for this was that different criteria were used to ground these similarities and the categorisations based upon them. Whilst each party recognised the similarities exploited by the other, they chose to ignore them, since those similarities were not salient for their personal underlying purpose or motive.

In the terminology of the explanation-based theory, it is necessary that similarity and categorisation be supported by a certain explanation or theory which justifies why certain similarities are chosen over and above others and so why those particular similarities comprise the basis of our categorisations. The explanation underlying Owen's similarity judgements involved his hatred of Lamarckism and transmutation theories in general, together with his antipathy towards Grant. All these factors caused him to exploit the mammalian similarity over and above the reptilian one. The explanations underlying his predecessors' judgements are less clear. However, I

hypothesized that they might have involved entrenchment and perhaps a passive and unconcerned adherence to transmutation.

A second (and related) point illustrated by Desmond's study which accords with the explanation-based view is that categorisation involves much more than simple attribute-matching, that it is only within the context of a particular theory or explanation that the choice of certain attributes over others makes sense. Owen took attributes such as size and certain bone peculiarities as relevant for placing dinosaurs in their own category and for claiming that they closely resembled pachydermal mammals. Again, his choosing these attributes rather than those chosen by his predecessors was dictated by his feelings towards Lamarckism and towards Grant – the explanation underlying his judgements.

Finally, in order to understand why Owen created the new taxonomic group of dinosaurs, we need much more than a list of conditions which defines that category and an illustration of the fact that these defining conditions are possessed by each member of the group. Rather, we need to understand Owen's motivations and reasons for doing what he did. It is important to see how his categorisation of the dinosaurs was related to his overall ideology, to his aims and commitments. It is only by taking all these factors into account that we can appreciate to the full the complexities which classification involves. Any reasonable account of classification must therefore take on board more general relational/explanatory factors, in addition to lists of defining conditions. This accords with a third major claim made by the explanation-based view – that our concepts are essentially relational and acquire much of their meaning through their interaction with other concepts and through the role that they play in the wider world context.

"Barnacle Larvae in the Nineteenth Century: A Case Study in Taxonomic Theory" – Mary P. Winsor.

Winsor's case study deals with the finding by one John Vaughan Thompson (army surgeon and field naturalist) that barnacles begin their lives as larvae. She discusses how this affected the classification of barnacles and shows how Thompson's discovery was utilised by different zoologists in different ways, indicating, "the active role theoretical considerations play in the process of zoological classification."[208]

[208] Winsor, 1969, p.309.

By 1823, Thompson had begun to suspect that many plankton were in fact none other than the young of species already known to zoologists. That spring, he caught a planktonic crustacean of a known genus (Zoea). He managed to keep it alive long enough that it began to moult, freeing five of its new legs. This was adequate to show that the new creature would have been very different from the original plankton; the structure of the limbs convinced Thompson that the genus Zoea was actually a larval crab. In 1826, he had better luck, at last capturing some "bivalved crustacea with stalked eyes"[209] and keeping them alive long enough that they moulted and became barnacles, thus adding further weight to his thoughts concerning plankton.

This was an important finding for two reasons. Firstly, it had previously been assumed that one of the factors separating the higher from the lower crustacea was that the lower underwent metamorphosis, whilst the higher did not. Absence of metamorphosis was also one characteristic sometimes used by taxonomists to separate the higher crustacea from insects. Secondly, it had previously been thought that barnacles did not belong to the class of crustacea at all. Linnaeus had grouped them with molluscs on the basis of their calcareous shells and lack of segmentation; in 1809 Lamarck had claimed that they in fact formed a transitional class (Cirripedia) between the annelids (worms with segmented bodies) and the molluscs. As a result of his discovery, Thompson was convinced that barnacles were neither molluscs or cirripedes, but belonged with the crustacea.

However, Thompson wanted more evidence for his view that the higher as well as lower crustacea underwent metamorphosis. Eventually, in 1827, he caught some female crabs along with their eggs (on the abdomen) and these eggs hatched. What emerged from the eggs were not baby crabs, but zoeae. Here was the evidence he needed. He then published his results and claimed that metamorphosis such as occurred with the common crab was a general principle among the higher crustacea (although he was aware that there were some exceptions to this rule). Thompson's research was repeated and accepted over the next few years, despite an initial rejection of his claims.[210]

In 1830, Thompson published his material about barnacles. He described the

[209] Winsor, 1969, p.295.

[210] Winsor tells us that his work was initially rejected for two reasons; firstly because of research by Heinrich Rathke which showed that the crayfish (classified as one of the higher crustacea) did not metamorphose and secondly because the general opinion was that metamorphosis would not occur in any other higher crustacea.

metamorphosis of his stalk-eyed crustacea into sessile (non-stalked) barnacles and hypothesized that stalked barnacles would have similar larvae to their sessile cousins. Later that year he was able to test this hypothesis when ships came in to Cork Harbour with stalked barnacles plus ova on their bases. However, the larvae which hatched were quite different from those of the sessile barnacles. He thus concluded that the two types were less closely related than he had supposed.[211] As Winsor tells us, "This was a reversal of opinion from his earlier memoir, and shows the great taxonomic weight which Thompson gave to the larval forms."[212]

Following Thompson's (correct) discovery that the young of barnacles exist in *larval* form, Winsor discusses how various eminent zoologists treated and made use of this new information. The thing to bear in mind here is that this new information about barnacles in no way clinched their classification. In fact, opinions as to how they ought to be classified varied widely. How an individual felt that they should be classified depended on his assessment of the importance or relevance of this information, which in turn reflected his distinctive theoretical commitments.

Henri Milne-Edwards apparently praised Thompson's work highly once he found that it tallied with work produced by Hermann Burmeister a year earlier. Milne-Edwards had previously shown himself willing to accept information from larval forms as indicative of taxonomic membership. In fact he had, since 1829, been formulating a theory which stressed the fundamental importance of embryology in determining class affinity. Winsor states that this theory must have impacted upon his classification of the Cirripedia as Crustacea when he revised his classification of the Crustacea in 1852.[213]

In 1835, Martin-Saint-Ange attempted to establish the 'correct' place of the Cirripedia class and relied hardly at all on Thompson's evidence in so-doing. He effaced the similarities which had earlier been claimed to exist between the cirripedes and molluscs, most significantly showing on the basis of a detailed anatomical study

[211] At this time, Thompson was unaware of the fact that barnacles in fact hatch out as what are now called nauplii and then later change into a different form now known as the cypris larva. So, in the 1826 case, Thompson had *actually* seen cypris larvae change into sessile barnacles, whilst in 1830, he saw nauplii hatch from the eggs of stalked barnacles. Since he was not aware of the two stages in the life of a young barnacle, he attributed the difference to some major discrepancy between the two forms of barnacle.

[212] Winsor, 1969, p.299.

[213] Prior to revision he had stated that perhaps, in the future, it might turn out to be necessary to classify the Cirripedia with the Crustacea, but that currently the scientific data was not adequate to make such a decision.

that earlier researchers had been wrong to think that barnacles were unsegmented. His study illustrated that barnacles share more similarities with the Crustacea than with any other class and that the major similarity they share with the annelids is that both are hermaphroditic. He thus placed the Cirripedia as a separate but transitional class between the Crustacea and the Annelida, believing that this was the 'natural' position for this class, forming the transition to the lowest of articulates (possessing joints or jointed segments).

Martin-Saint-Ange considered one of the advantages of his classification to be that it better corresponded to Ampère's zoological system. Winsor tells us that Ampère, "had arranged the animal kingdom into two parallel series, each of which ran from the simple to the complex, and which was symmetrical...each group in some way corresponded to the group at the same level on the other series."[214] Martin-Saint-Ange apparently argued that grouping the cirripedes with the molluscs had not fitted in with this system, but according to his re-classification, the cirripedes were now opposite the cephalopods (certain types of marine molluscs, characterised by well-developed head and eyes and a ring of sucker-bearing tentacles, such as octopuses, squid and cuttlefish), the correspondence consisting in possession of a soft body covered by a shell. This was supposed to iron out the previous irregularity.

James Dwight Dana dismissed systems such as that of Ampère. In general, he followed Milne-Edwards' judgements, but he was also very interested in the grade of animals; whether they were superior or were degradations. He based his ideas on the heuristic that superior animal grades show a greater centralisation, meaning that they exhibit a victory of animal over vegetable forces. The cirripedes were immobile and so Dana argued that here the vegetable forces had become so entrenched that they had overcome any intrinsic motive energy which the cirripedes might have possessed. However, he was positive that Thompson's evidence proved that they were true crustacea and so was unconcerned by the fact that the cirripedes are immobile, whilst crustacea are very definitely mobile. Rather, following his own theory, he believed that the cirripedes simply represented a lower grade within the Crustacea class.

Richard Owen was well acquainted with the work of Thompson and was of the opinion that larval forms were very important for determining an animal's taxonomic position. However, he did not agree with Thompson's conclusion that cirripedes were

[214] Winsor, 1969, p.301.

crustacea. He considered that an increase in powers of movement was indicative of progress in an animal. He claimed that although the cirripedes possessed jointed appendages, these were used primarily for food gathering, rather than for the sake of locomotion. He therefore opted to place the Cirripedia as a separate class in between the Crustacea and the Annelida. Many of his reasons were the same as those of Martin-Saint-Ange. Owen persisted with this view even after Darwin had published his monograph in 1854, stating that the cirripedes belonged in the Crustacea class (monographs were usually taken as definitive). He repeated the previously stated reasons for his classification and added that although he considered the larval cirripedes to be crustacean, once they lost their mobility as adults, they could not be considered members of the Crustacea class.

Darwin apparently felt no need to explain in any detail his belief that cirripedes were crustacea. He felt that the adult morphology of barnacles which had been discussed by Martin-Saint-Ange should be adequate to transform the Cirripedia into a sub-class of the Crustacea (contrary to Martin-Saint-Ange's decision). However, he did illustrate that some of the non-crustacean features which had been attributed to barnacles were in fact mistaken or based on insufficient knowledge. For instance, he showed that not all barnacles are hermaphroditic. He also did a little work on larval morphology, illustrating the similarities between cirripedian and crustacean larvae, commenting that members of one group are often more alike as young than as adults.

Winsor concludes:

> "If cirripedes were to be included in the Crustacea, zoologists would have to alter their concept of crustacea as being locomotive bisexual animals...An individual zoologist's decision was often clearly related to his beliefs about what a taxonomic system should look like. For example, Martin-Saint-Ange and Richard Owen subscribed to the traditional view that animals could be arranged in linear series, and viewed the Cirripedia as a link between two larger classes. Milne-Edwards and Dana...saw in nature irregular groupings around conceptual types...interpretations of the importance of the larval form both depended upon and influenced general theories about the taxonomic significance of embryology. The variety of uses to which zoologists put Thompson's discovery of larval barnacles reflects the active role theoretical

considerations play in the process of zoological classification."[215]

Evaluation of Winsor's Case Study.

Winsor's case study helps to illustrate the intricate complexities surrounding the classification of entities in the natural world. It is interesting in that it shows how different theoretical commitments cause scientists to treat the same piece of information in differing ways, or with differing degrees of importance. More specifically, some scientists considered embryology to be very important for determining classification, hence they took larval form to be indicative of membership in the Crustacea class. Some considered embryology to be less important for such matters, whilst others took embryology to be important, but in the case of barnacles, saw other factors as overriding the significance of the larval form.

Factors which Influenced the Classification of Barnacles.

Prior to Thompson's discovery, the barnacles had been classed differently, according to different individuals. Linnaeus saw them as molluscs on the basis of their calcareous shell and apparent lack of segmentation, whilst Lamarck felt they needed a class to themselves, albeit transitional, then Cuvier denied that there could be any intermediate between annelids and molluscs, coming to the same classificatory conclusion as Linnaeus. We can already, at this stage, see the different interests at play. Individual scientists are taking different factors to be salient or important for classifying the barnacles. For Linnaeus, certain physical facts about barnacles were to be taken as salient for classification. Lamarck presumably latched on to other features of barnacles, by which he saw them as related to both the annelids and the molluscs. He took this mixture of features to be salient and judged it sufficient for the formation of an independent class, which, nonetheless, could be seen as the link between two other classes with which the cirripedes shared many features. For Cuvier, however, such a link was impossible. He brought this prior commitment to his work and on the strength of it, chose to class the barnacles with the molluscs. Thus even prior to Thompson's discovery, classification of barnacles was anything but a simple affair. Different scientists approached their work with prior commitments, theories and beliefs, all of which resulted in conflicting accounts of where barnacles should be placed taxonomically.

When Thompson observed his larvae transform themselves into barnacles, he was

[215] Winsor, 1969, p.309.

convinced that this proved beyond any doubt that they were true crustacea, rather than molluscs or a transitional class. Prior to Thompson, no one had considered that barnacles might be members of the Crustacea, since no one knew that they began their lives as larvae and underwent metamorphosis; classification of barnacles was determined purely on the basis of the adult form. As Winsor herself points out, the characteristics which the taxonomists were looking at – the calcareous shell, hermaphroditism and the visceral anatomy (the large internal organs of the body) – are all left when the barnacle has moulted. However, if the exuviae (moulted skin) is examined, it is apparent that the cirripedes moult, that their limbs are jointed and that their jaws are like those of many crustacea. The question then becomes one of *which* characteristics are to be taken as salient for matters of classification.

Clearly, Thompson saw *larval* form as overriding for judgements about classification. Since many of the larval characteristics of barnacles were similar to those of the crustacea and since it was known that some of the lower crustacea (and later accepted that some of the higher crustacea) undergo a change of form, Thompson concluded that the barnacles fell firmly in the Crustacea class. For him, embryology and embryonic features were fundamental for classifying entities in the natural world. This represents a prior commitment or theory which he brought to his work and which influenced his judgements, thus ensuring that certain features would be salient for him over and above others in matters of classification. My point is supported by Thompson's confusion of the stalked and sessile barnacles. He hypothesized that the stalked barnacles would begin life as larvae very similar to those of their sessile counterparts. However, on the basis of the subsequent apparent *difference* of larval form (see footnote 168), he was prepared to renege on his hypothesis and conclude that the stalked and sessile varieties must be less closely related than he had previously supposed. He took larval form to be so important for classification that he was prepared to say that two specimens which shared a very large number of adult features in common were, in fact, not closely related simply because, as larvae, they appeared to be considerably different.

Milne-Edwards set great store by Thompson's discovery of barnacle larvae and later came to the conclusion that the cirripedes were undoubtedly crustacea. However, we learn that he had been developing a theory which emphasised embryology as being highly important for revealing taxonomic affinity. This presumably acted as a prior theoretical commitment which he brought to his work and which ensured that the

embryonic features of barnacles would strike him as the salient ones for matters of classification. His prior commitments therefore led him to concentrate on the relevance of the larval form for classification.

Despite the certainty of Thompson and Milne-Edwards that barnacles were crustacea and despite their reliance on what they saw to be the fundamental nature of the larval form for classification, we learn that other scientists had different ideas. Martin-Saint-Ange, for instance, carried out a very careful anatomical survey of the barnacle and paid very little attention to Thompson's evidence. He noted that the cirripedes shared more in common with the Crustacea than with any other class, yet he still grouped them in a class of their own, transitional between the Crustacea and Annelida. Why might this have been? Martin-Saint-Ange claimed that the one thing which the cirripedes had in common with the annelids was hermaphroditism. Perhaps this is why he placed the cirripedes between these two classes, rather than with the crustacea. But why did he consider the hermaphroditism to be of such weight that it drew the cirripedes out of the Crustacea and into their own distinct class? By doing this, Martin-Saint-Ange was able to rectify what he saw as an irregularity in Ampère's zoological system. Winsor tells us that systems such as that of Ampère which aimed to, "find a logical, regular pattern in taxonomic systems"[216] were highly popular at this time. The best known system was that of William S. MacLeay and Winsor states that, "A considerable number of zoologists, if they did not embrace MacLeay's system, did sympathise with his attempt..."[217] I suspect that Martin-Saint-Ange, tuned to the climate of the time, believed that there was some kind of regular pattern to be found in taxonomy. He considered that by keeping the Cirripedia as a class in its own right, one irregularity in Ampère's system could be ironed out, thus providing more grist for the mill in the attempt to show that such systems were well-founded. This would have acted as a prior belief, commitment or theory which he brought to his work. Commitment to the truth of such a system ensured that characteristics which would in some way provide support for the system would be particularly salient for him. This transformed the hermaphroditism of the cirripedes into a highly salient and important feature, which enabled him to place them as a transitional class between the Crustacea and Annelida.

James Dwight Dana, on the other hand, rejected the zoological systems of scientists such as Ampère and MacLeay, considering that they succeeded only in imposing an

[216] Winsor, 1969, p.301.
[217] Winsor, 1969, p.301.

artificial arrangement on nature. He was therefore unlikely to agree with Martin-Saint-Ange that the cirripedes formed a transitional class; he lacked the motivations and theoretical commitments which would drive such a claim. Winsor tells us that he generally accepted the judgements and conclusions of Milne-Edwards, agreeing that Thompson's discovery placed the barnacles unequivocally in the class of Crustacea. It is interesting, however, to compare his views on the immobility of the adult barnacles with those of Richard Owen (below). For Owen, mobility was crucial to the Crustacea class and for this reason, the cirripedes could not be crustacea. For Dana, by contrast, mobility (or the lack of it) were indicative of superior or lesser grade within a class. It did not concern him that crustacea are characteristically mobile, whilst adult barnacles are not. Following Milne-Edwards, he felt that the larval form provided more than enough evidence that the cirripedes were true crustacea. The fact that adult barnacles lost their mobility was simply incorporated into his theory of centralisation – vegetable forces were dominant in these animals, hence they represented a lower grade within the Crustacea class. For Dana, embryology was apparently dominant for revealing taxonomic affinities. His own theory of centralisation, whilst relevant for determining grade or level within a class, was not relevant for determining class membership.

Richard Owen, it seems, was well acquainted with Thompson's work and apparently set high store by the larval forms discovered, yet he still refused to count the cirripedes as crustacea. As has already been explained, Owen saw locomotion as a fundamental characteristic. As far as he was concerned, increase in locomotive power meant progress. Those creatures already classed as crustacea had considerable powers of spontaneous motion. Not so with the cirripedes which, Owen argued, moved only for the sake of finding food. Although he recognised that barnacle *larvae* were free moving and acknowledged the importance of this discovery, still he insisted that they must not be classed in the Crustacea since, as adults, they lost this faculty of spontaneous locomotion. Thus, despite the fact that Owen considered Thompson's discovery of larval forms to be very important, there were still, for him, other more crucial factors which served to override this one in matters of taxonomy. Owen clearly brought his belief in the importance of locomotion to bear very strongly on his work; this effectively served as a prior theory or commitment which influenced his judgements concerning taxonomic position. He therefore argued that since the adult cirripedes lost their powers of movement, they were clearly not as advanced as the fully mobile crustacea and so their proper position lay between this

class and the Annelida; a separate transitional class leading from the jointed articulates to the mobile crustacea.

> "We make the easiest and most natural transition from the lower forms of Articulata [the annelids] to the Crustaceous class, by passing from it to the larval state, – which I have argued to be the typical one [that is, revealing the type] – of the...*Cirripedia*; in which view we may regard the *Crustacea* as representing those larvae on a gigantic scale, and so retaining the typical character with the faculty of motion."[218]

Darwin was convinced that the cirripedes were to be included in the class of Crustacea. He believed that Martin-Saint-Ange's morphological studies should have been adequate to justify classing them in this way. He did not believe in the reality of a system such as Ampère's and hence was not driven by such commitments to make the cirripedes a class unto themselves. Darwin, however, had his own commitment which was pulling him the other way and which he brought to bear on his taxonomic work. This was his own theory of evolution. He apparently saw Thompson's discovery and embryology in general as being very important. As Winsor informs us, "The larvae were important to him primarily as a means of determining the homologies of the parts of cirripedes with other crustacea."[219] In other words, Darwin was keen to show that the cirripedes shared a common evolutionary origin with members of the Crustacea. If he was able to show this, he would be bound, by his own theory, to consider the cirripedes as belonging with the crustacea. Shared evolutionary origin was, for him, the most salient fact for matters of classification.

Milne-Edwards had outlined a theory which stated that crustacea are made up of a series of homologous segments and which suggested an archetypal number of twenty-one segments for the average crustacean. Darwin had taken up this idea and applied it to the larval cirripedes, announcing that the relevant number of segments here was seventeen. This was close enough to twenty-one that it was quite feasible, on the strength of this calculation, to include the cirripedes in the Crustacea. The principle behind these calculations was that larval forms are more likely to retain ancestral characteristics than adult forms. In other words, members of one group will often be more alike when young than as adults.

---

[218] Owen quoted in Winsor, 1969, p.304.
[219] Winsor, 1969, p.305.

Darwin had a prior commitment to his own theory of evolution. This led him to believe that evolution played a vital part in fixing taxonomic position; those creatures which are descended from a common ancestor should be grouped together. Part of the theory declared that larval forms are the most likely to retain ancestral traits. Given these facts, we would expect Darwin to consider larval form and embryology in general as highly relevant for classification. It is the larval form of the cirripedes which is crustacean in character; once they reach adulthood, the resemblance is vastly reduced. Hence Darwin was bound to class the cirripedes with the crustacea. For him, characteristics of the larval form were conclusive, picked out as salient by his prior commitment to the theory and consequences of evolution.

The Inherent Messiness of Science.

Winsor's study illustrates, once again, that science is not the cut-and-dried affair which, I believe, people assume it to be. She shows that different scientists held different opinions regarding how barnacles ought to be classified and that these alternative opinions were justified by different theories and beliefs. Furthermore, it is clear that the alternative classifications rest upon possession of real properties or characteristics and so each reflects a real or objective aspect of the natural world. Her study helps to dispel what, I think, are widely held beliefs that there is only one scientific answer and that this answer is a reflection of the world's 'true' nature, which leads to the notion that, once the question has been settled 'correctly', there should be no further dispute or discrepancy in decision between individual scientists.

Categorisation as More than the Way the World Is.

Winsor's case study provides us with good evidence against the picture of science as reflecting a unique taxonomy which is inherent in the natural world. We again have evidence which illustrates the *complexity* of the natural world, with different patterns of similarity cutting across it. The cirripedes were said to bear similarities to the crustacea, to the annelids and to the molluscs (until Martin-Saint-Ange denied this). With this kind of complexity, is is unlikely that there will be only one available answer and, as we have seen, there are various ways in which the cirripedes could be (and were) classified. It became a case of picking and choosing between similarities which really exist in the natural world, of determining which ones were to be taken as salient. Inevitably, in such a situation, conflicting answers emerge. It is not the case that one of these answers is correct because it truly reflects the way nature is and that all the others are wrong. Rather, each answer reflects a different way of

looking at the same thing; it focuses on one specific set of similarities. There is no denying that objective and real similarities exist between the cirripedes and both the annelids and crustacea. The cirripedes and crustacea both exhibit larval forms with similar jaws and jointed limbs. The cirripedes and annelids both have segmented bodies and at least some of the cirripedes also share the similarity of being hermaphroditic. Since the natural world displays these different patterns of similarity, we must go beyond the the world itself to the motivations of the taxonomist in order to explain how and why different scientists pick out different similarities as salient.

Human Input for Categorisation: Aims, Theories and Commitments.

Once we have acknowledged that the world itself is able to support differing classifications on the strength of alternative patterns of similarities, we are in a position to recognise the substantial amount of human input involved in classification. As I have shown in detail above, the way individual scientists chose to classify the cirripedes was dependent upon their prior commitments or beliefs in the truth or relevance of certain theories. These prior commitments therefore determined which similarities were salient for them. This was not simply a question of how to categorise the barnacles, but was also a question about the relevance of embryology to taxonomy as a whole. It seems to have been universally admitted (with the exception of Martin-Saint-Ange, perhaps) that Thompson's discovery of barnacle larvae was very important and had potential relevance for taxonomy. This suggests that people were generally of the opinion that embryology was a relevant factor for taxonomy. However, this was not the whole story. People might see embryology as important, but whether they saw it as the *conclusive* factor in matters of taxonomy would depend upon the other beliefs or commitments which they held. This, in turn, affected their decision concerning the classification of barnacles in particular. And so human input in the form of prior aims, theories and commitments played a highly important part in barnacle classification.

The Blurring of the Line between Metaphysics and Epistemology.

Winsor's study provides us with more evidence in support of the claim that the metaphysics/epistemology distinction is blurred and that, as far as matters of categorisation go, it makes no sense to separate the two. It has been illustrated that in the case of zoological taxonomy, the natural world clearly does not exhibit one inherent taxonomy. Rather, the natural world is able to support alternative and conflicting classification of the entities within it. In this case, we cannot say that

scientific classification is simply a matter of metaphysics, of the way the world is. The world is a highly complex and intricate place, exhibiting diverse patterns of similarities and affinities which might be chosen as the basis for classification. What this means is that epistemology must play a crucial role in scientific taxonomy. Scientists must (and do) approach their work with prior commitments to certain beliefs or theories. Such commitments enable them to pick out certain similarities as particularly salient for taxonomy. Scientists who have different prior commitments may well pick out different similarities as salient and hence produce conflicting classifications. This occurred with the barnacles; everyone considered Thompson's discovery of barnacle larvae to be highly important, but different individuals made use of this information in different ways, dependent upon their allegiances to certain theories and on the strength of their beliefs in the importance of embryology in general for revealing taxonomic affinities. We can see, then, that when scientists classify entities in the natural world, their decisions represent a fusion of metaphysics and epistemology; they base their decisions on certain real similarities existing in the environment (metaphysics), yet their choice of certain similarities as the relevant ones will be guided by their prior commitments (epistemology).

The Problem with Essentialism.

It seems that those philosophers who want to claim that certain unique groupings (natural kinds) exist in nature by virtue of the fact that each and every member of the same group possesses some kind of essence in common with all other members and that it is the job of science to discover this natural taxonomy together with its essential basis and correctly mirror it, have again been shown to be wrong. We have already seen that nature is capable of supporting more than one classification and that which one is accepted by which scientist will depend on the commitments of that scientist. Thus any story about one unique taxonomy based on possession of certain essences must be a misguided one. Both the grouping and the definition of a particular group have been shown to depend as much on epistemology as on metaphysics. Thus Winsor states, "If cirripedes were to be included in the Crustacea, zoologists would have to alter their concept of crustacea as being locomotive bisexual animals...An individual zoologist's decision was often clearly related to his beliefs about what a taxonomic system should look like...interpretations of the importance of the larval form both depended upon and influenced general theories about the taxonomic significance of embryology."[220]

[220] Winsor, 1969, p.309.

Support for an Explanation-Based Model of Categorisation.

This case study provides us with more good evidence for the claim that the explanation-based account of categorisation applies as much to scientific as to lay categorisation.

Winsor's study highlighted acutely the very significant role which theory plays in zoological taxonomy in general and played with reference to the classification of barnacles in the nineteenth century in particular. We have seen that how an individual scientist chose to classify the barnacles depended very heavily on his prior commitment to certain theories or beliefs. Thompson, Milne-Edwards and Dana all strongly believed that embryology was vital for fixing taxonomic affinity, hence they concentrated on the crustacean features of the barnacle larvae, classifying the cirripedes with the crustacea. It was hypothesized that Martin-Saint-Ange was heavily influenced by his belief in Ampère's system, which led him to place the cirripedes in their own class between the Crustacea and the Annelida. Owen was influenced by his overriding belief in the fact that locomotion represented progress, which meant that he refused to class the barnacles in the Crustacea. Darwin believed that descent from a common ancestor should be reflected in taxonomy and that the larval form was more likely to retain ancestral traits, which led him to place the barnacles in the Crustacea. The details of how and why these different scientists classed the barnacles as they did closely tallies with the basic claim of the explanation-based account – that there must be some kind of theory or explanation which underlies and so gives sense to the similarity judgements and categorisations which human beings make.

Since each scientist approached his work with and was guided by a particular theory or belief, that belief or theory determined which similarities he took to be salient for classification. Those scientists who, for whatever reason, took barnacle larvae as indicative of class membership saw their crustacean features as highly salient, as being the similarities which count for classification. For those scientists who saw larval form as less important, the crustacean characteristics of the barnacle larvae lost their overriding saliency. These scientists considered characteristics of the adult form to be salient as well and barnacles were seen as sharing characteristics with both the crustacea and the annelids. In this way, different scientific theories or explanations can be seen to highlight different shared properties or similarities as salient for matters of classification. This reflects the explanation-based notion that it makes little

sense to say that two things are similar *per se,* since they can be both similar and dissimilar in numerous ways. Rather, some sort of theory or explanation is required to account for *why* certain properties are taken as relevant over and above others and to provide a backdrop against which certain similarities become salient for the purposes of classification. The data from Winsor also illustrates the explanation-based claim that categorisation comprises more than simple attribute matching – it also involves relations between attributes and explanations of why some attributes are judged to be more relevant than others.

A final point made by the explanation-based theory is that concepts do not have meaning in and of themselves. Rather, they gain a large amount of their meaning from the role that they play in the wider world, from the ways in which they are related to or associated with other concepts. This kind of thinking is applicable to Winsor's case study. In order to understand how and why individual scientists classified barnacles as they did, it has been necessary to consider their wider beliefs and motivations. Their taxonomic decisions only make sense in the wider context; their decisions were governed by factors extraneous to the barnacles themselves. To fail to take these factors into account is to fail to see the complete picture.

Some More Assorted Examples.

Three further examples of classification are summarised very briefly below. All three illustrate the main points which I have argued above in relation to the case studies – that categorisation comprises more than the way the world is, that there is a degree of human input to categorisation and so that classification represents a combination of metaphysics and epistemology. The first example deals with the different ways in which diseases have been defined through history, whilst the second and third are more theoretical in nature.

Ludwik Fleck's book, *Genesis and Development of a Scientific Fact*, deals with the historical development which led up to the modern concept of syphilis. Following modern methods, diseases are defined etiologically. That is, they are defined by their cause. Fleck points out that this is not, however, the only logically possible way of defining diseases – they might be defined symptomatically, for instance. Using the etiological concept of disease, there is no way in which syphilis and leprosy can be said to be related, since syphilis is a spirochaetosis[221] and leprosy a disease caused by

[221] This is a disease caused by any bacterium of the order Spirochaetales.

the specific bacillus. However, Fleck emphasises that if diseases are defined symptomatically, then the two could be said to be closely related, since their symptoms are sufficiently similar. He therefore claims that the historical statement, "The French pox, or syphilis, or the carnal scourge, which is the consequence of contagious and leprous affections of the genitals, is a daughter of leprosy and can in certain circumstances in turn become the mother of leprosy."[222] although meaningless in connection with the etiological disease concept, in fact makes perfect sense in relation to the symptomatic disease concept.

David Bloor in his paper, "Durkheim and Mauss Revisited: Classification and the Sociology of Knowledge", pictures a very simple conceptual system, whereby 'fish' are said to live in the sea, 'birds' to fly in the air and 'animals' to live on the land. Part of the class of animals suckle their young and these are known as 'mammals'. Bloor then imagines that the users of the system come across a new beast – the whale – which looks like a fish, yet suckles it young. At this point a decision has to be made – either we have to say that not only animals but also some fish suckle their young, or we have to say that not all mammals live on the land. Bloor's point is, "Resemblance alone pulls us in opposite directions, so the choice cannot be resolved by experience. Retaining either law will do justice to what is known and restore an acceptable form of correspondence with the world. And, "...the functioning of a name or predicate cannot be explained fully by similarity relations but depends on the laws into which they enter. Conversely, the laws will depend on conventions about the boundaries of the classes they relate. This reciprocal dependence: the discretion it allows, and the choices it forces upon us, are completely general phenomena."[223]

N.R. Hanson in his book, *Patterns of Discovery,* imagines two microbiologists looking at an amoeba. He claims that one might see it as a one-celled animal whilst the other might see it as a non-celled animal. The first microbiologist sees the similarities which the amoeba bears to all other cells – it has a nucleus, cell wall, cytoplasm and so on. The other sees the similarities which the amoeba bears to all other animals – it is mobile, ingests and digests food, excretes and reproduces.

[222] Simon, quoted in Fleck, 1935/1979, pp.121-122.
[223] Bloor, 1982, p.274.

Conclusions.

Through summarising and evaluating the three case studies, I have illustrated six interrelated points regarding scientific classification. These are that:

- Scientific classification is not a straightforward, simple or clear-cut affair.
- Scientific classification involves much more than charting the way the world is.
- Scientific classification involves human input in the form of aims, theories and commitments.
- The line between metaphysics and epistemology is blurred.
- There are problems with applying the objectivist doctrine of essentialism to scientific classification.
- Scientific case studies lend support to my hypothesis that the explanation-based model of categorisation applies as much to classification by scientific experts as to classification by laypeople.

Clearly, these six point are closely related and are, to a certain extent, mutually self-supporting. Primarily, the case studies have thrown doubt on the conception of scientific classification as producing the one final answer regarding how entities in the natural world should be grouped and about which all scientists are in agreement. This leaves room for the notion that there are differing and conflicting theories regarding what properties or features ought to count as relevant for matters of classification, which sometimes lead to competing classifications – either regarding the defining properties of a particular class, or regarding which entities are to be included within a particular class, or both. We have seen, furthermore, that these competing classifications are all based upon real or objective properties which entities exhibit and upon real and objective similarities or regularities which entities share. Competing classifications reflect different real aspects of those entities and for this reason, the world lends support to each of the competitors – it does not give preference to one over the other. Since the world is neutral in this way, classification must involve more than a reflection of the way the world is. Since the world 'is' in more than one way, some form of human input is required to select one or more of these ways and to make them the basis of a (system of) classification. Categorisation

must therefore involve a mixture of metaphysics (one of the ways in which the world is) and epistemology (human theorising or belief which serves to make a choice between different ways in which the world is). If this is so, then philosophical essentialism is not a suitable model for scientific categorisation, since it claims that there is a unique metaphysical division of the natural world into kinds and their respective essences and that it is the role of science to uncover this division, untainted by any kind of theoretical or 'subjective' commitment. Finally, all of these points combine together to lend support to the claim that the explanation-based model applies to scientific classification. This is because they highlight the notions that entities can be similar in a number of different ways, that an underlying theory or explanation is therefore required to account for why certain attributes and similarities are chosen over and above others for matters of categorisation, that classification therefore involves more than attribute-matching and that concepts are relational entities rather than isolated entities which are characterised in terms of a list of defining or necessary and sufficient conditions.

It is clear that these six points which I have argued that the case studies illustrate deal a devastating blow to the doctrine of Objectivism, as characterised in Chapter Two. In particular, they militate against the doctrine of essentialism and the objectivist conception of natural kinds, as recently revived by Putnam and Kripke. They also militate against Rey's objectivist position, especially his insistence upon the existence of a rigid division between metaphysics and epistemology, as discussed in Chapter Five. It is important to note, however, that no blow has been dealt to objectivity or realism. Nowhere is it denied that humans are in touch with the world outside of themselves, nor is it claimed that they construct the world around them. Rather, scientists make use of and choose between real and objective properties, similarities and attributes when classifying entities in the natural world.

My evaluation of the case studies also lends strong support to the (later) Putnam's internal realism and to Johnson's experiential realism, as summarised in Chapter Two. Specifically, we have learnt that even in relation to *scientific* classification, the notion of a God's Eye view does not make sense, since there are a number of different aspects of and so different ways of looking at the world. We have also seen that, as a result of this, some kind of human input in the form of choice between attributes and similarities is essential in order to get classification up and running. Therefore, it does not make sense to construe human beings as rational agents who

are in some way detached from and external to the environment. Rather, they must be seen as a dynamic part of that environment, both affecting and being affected by their surroundings.

Chapter 7
Philosophical Arguments.

Introduction.

In this chapter, I want to consider some philosophical arguments which cover similar ground to my own arguments and conclusions discussed in the previous chapter. Typically, these philosophical arguments are reactions against the rigid realism espoused by Putnam and Kripke in relation to natural kinds and essentialism. In particular, they make the point that division of the world into natural kinds involves an element of human choice or decision and so is not solely determined by nature itself. The proponents of these arguments can therefore be be seen as rejecting Objectivism in relation to natural kinds as explained in Chapter Two.

Keith Donnellan.

In "Kripke and Putnam on Natural Kind Terms", Donnellan examines Kripke and Putnam's views on reference. In particular, he wants to argue that the extension of their theories from singular terms to general terms does not work in the straightforward manner which they suppose. He believes that a theory of natural kind terms will require a different and more extensive treatment.

Donnellan isolates the major problem with Putnam and Kripke's account of natural kinds terms as being the following. Proper names are singular and kind terms are general. Thus if two singular terms are both rigid designators and both have the same extension, then both terms will designate a single entity in the actual world and hence must designate the same entity in all possible worlds. This formula does not, however, transfer to natural kind terms. Just because two kind terms have the same extension in the actual world, it does not mean that that the two terms therefore name the same kind. Co-extension does not necessarily indicate identical kinds. The example employed by Donnellan is the well-worn one of 'hearted thing' and 'livered thing'. As a matter of fact, in the actual world, all things which have a heart also have a liver – they are co-extensive. However, it does not therefore follow that 'things

with hearts' and 'things with livers' are one and the same kind – they are not. They are two separate kinds which happen to be co-extensive.

Putnam and Kripke are apparently aware of this difficulty. They realise that they are going to have to provide something more in order for their theory of natural kinds to work. What they do (and here Donnellan concentrates on Putnam's account) is introduce the idea of important physical properties which are discoverable by science and which therefore provide the true definition and demarcation of the kind in question. Donnellan quotes Putnam here:

> "To be water, for example, is to bear the relation $same_L$ [same liquid] to certain things."[224]

> "*x* bears the relation $same_L$ to *y* just in case (1) *x* and *y* are both liquids, and (2) *x* and *y* agree in important physical properties."[225]

Of course, the point that Putnam is trying to convey here, using water as a specific example, is that nature divides inherently into certain (natural) kinds and these kinds are underlain by possession of certain fundamental defining properties. Once these properties have been discovered, there is one and only one way in which we can correctly divide up the natural world and this is the way which reflects these (natural) kinds and their defining properties.

Donnellan, however, is sceptical that this account works quite as conclusively as Putnam would like to claim. He illustrates the gaps which he sees in Putnam's account by means of a thought experiment.

We are asked to imagine two cultures, one on Earth (and so identical with our own culture) and one on Twin Earth which, up to a certain point in their respective histories, are absolutely alike in every possible way. In particular, their languages are identical and at a certain point, they both developed identical sophisticated sciences and so embraced scientific viewpoints of the world. They both developed atomic theory. They therefore both have the concept of an atom as having a nucleus which comprises positively charged particles (protons) and neutrally charged particles (neutrons). They both have the concept of elements as non-compound substances whose atoms have a particular number of protons in the nucleus and they both call

[224] Putnam, 1975, pp.238-9.
[225] Putnam, 1975, p.239.

this number atomic number. They both have the concept of isotopes as being individuated by the combined number of protons and neutrons in the nucleus of the atom and they both call this the isotope number. They are both aware that one and the same element usually has several different isotopes and they both give separate names to each isotope of the element.[226]

Donnellan continues the story – on Twin Earth, it is generally the case that one of the isotopes of a particular element comprises the bulk of the naturally occurring element, whilst the remaining isotopes tend to be pretty rare. He indicates that he does not know whether this is actually the case on Earth, but that the point of this move is just to increase the psychological plausibility of the story – it is not, in fact, necessary for the story to carry through. Different isotopes of the same element display important differences in behaviour. For instance, some isotopes are radioactive, whilst others of the same element are not; some isotopes are unstable and break down into isotopes of other elements, whilst others of the same element do not. Of course, it will also be significant which *element* you are dealing with, particularly in relation to chemical reactions. The point which Donnellan is trying to get across here is that it may not be at all clear whether atomic number or isotope number is the more fundamental hidden property so far as the substances which we call the 'chemical elements' are concerned.

The details of this story are then applied to specific examples. Donnellan claims that it is not implausible that on Twin Earth 'gold' is used to label the isotope which makes up the bulk of the element with atomic number 79, rather than simply the element with atomic number 79, as on Earth. The Twin Earthians would then view the rarer isotopes as not being true gold, even though they are very similar to gold in certain respects whilst the Earthians make no such distinction.

The story can be extended beyond the chemical elements. On Earth, the term 'water' is used to refer to what we call 'heavy water' (deuterium or tritium oxide) as well as to protium oxide. However, we can imagine that on Twin Earth, water is identified solely with protium oxide. So-called 'heavy water' is considered by the Twin Earthian scientists to not really be water, since it is not composed of H_2O molecules, but of T_2O or D_2O molecules.[227]

[226] Donnellan points out that in actuality, isotopes are distinguished by a combination of atomic and isotope number, but for the sake of simplicity, he assumes that they are distinguished only by isotope number. Either way, the points which he goes on to make would remain the same.

[227] See Zemach, 1976 and Mellor, 1977 for arguments along similar lines.

This kind of situation has some interesting consequences which are noted by Donnellan. We can see that on the kind of account provided by Putnam, it is possible that in two cultures with exactly the same linguistic and scientific background, a particular kind term could have a different extension in each culture. Yet this is certainly a possibility which Putnam would wish to avoid. For him, natural kinds are just that – natural – they represent the inherent structure of nature and so by definition have only one possible extension. It begins to look, then, as if natural kinds must involve a little more than simple reflection of nature, since it seems that, if Donnellan is right, nature is in fact ambiguous and not decisive, as Putnam's story requires.

Secondly, as Donnellan points out, his story leads to a difference in the truth value of certain sentences, dependent on whether one is an Earthian or Twin Earthian. For instance, the sentence, "Some gold has isotope number x and some has isotope number y", will be true for Earthians, but false for Twin Earthians. The sentence, "Gold has atomic number 79", will be true for Twin Earthians, so long as the word "has" does not involve any notion of identity but straightforwardly true (by identity) for Earthians. The sentence, "Gold is identical with the element with atomic number 79", will be true for Earthians and false for Twin Earthians.

Donnellan concludes that we cannot agree with Putnam that it is obvious that natural kind terms have the same extension before and after scientific discovery of their hidden properties. Which way scientists choose to interpret and apply these discoveries will affect the extension of those terms. To use Donnellan's example – suppose Locke had had a ring made out of one of the rarer isotopes of what we call gold. He would have claimed that this ring was gold on the basis of its surface properties. Post-scientific discovery, the statement that this ring was gold would be correct for Earthians, but incorrect for Twin Earthians. On Twin Earth, the extension of 'gold' has changed *as a result of* scientific discovery due to the fact that isotope and not atomic number was taken as definitive of the natural kind. Donnellan concludes that if we go part of the way with Putnam, we are forced to, "...admit nature does not fully determine the extension of vernacular natural kind terms and science is not wholly responsible for discovering their true extensions."[228]

[228] Donnellan, 1983, p.104.

Analysis.

It seems that Keith Donnellan's work can be analysed in much the same way as I analysed the case studies in the previous chapter. Donnellan's remarks and thought experiment seem to provide support for the kind of view which I am trying to put forward in opposition to the Objectivism entertained by philosophers such as Putnam and Rey. Furthermore, all the case studies in the previous chapter concerned species in botany, biology and zoology, yet Donnellan's account shows how similar arguments can be applied to other frequently cited examples of natural kinds, such as the chemical elements.

Categorisation as More than the Way the World Is.

Donnellan's thought experiment provides us with a good example of how our division of the natural world into kinds involves more than a simple reflection of the inherent structure of the world.

Putnam is of the opinion that once science uncovers the important hidden properties of entities and substances, it will thereby have discovered the basis for the world's division into kinds. In other words, he sees this division as immanent in nature itself and it is only by following this inherent division that we can come up with a correct classification of the natural world.

Donnellan's thought experiment, however, indicates that nature is not as decisive as Putnam wants to claim. The problem, of course, centres around the isolation of *the* important property or properties. Often, scientists will discover *more than one* important hidden property. The problem then becomes one of deciding which property is the important one for matters of categorisation or division into natural kinds. This is a crucial issue, since which property is taken to be the important or relevant one will affect our division of the natural world – different properties result in different natural kinds.

The first major point which Donnellan attempts to get across in his Earth/Twin Earth story is that both atomic number and isotope number are important physical hidden properties, in the sense that Putnam's account appears to require – they are both properties associated with the microstructure of substances. Each property can be held accountable for certain uniform reactions which take place within or between particular substances. In other words, it is arguable that both atomic and isotope

number are equally good candidates for being *the* important property, the one which is relevant for dividing up the natural world and so determining natural kind membership. The natural world seems to be ambiguous here – it does not come pre-packaged with only one preferred property for natural kind membership.

It is because the natural world is ambiguous that Donnellan is able to construct his Earth/Twin Earth example. Even given two cultures with exactly the same language and scientific knowledge of the world, we are still left with natural kind ambiguity. As a matter of fact, scientists on Earth have taken atomic number as the relevant property for delineating the natural kind gold, but it is feasible that they might instead have chosen isotope number as the appropriately relevant property, as the scientists on Twin Earth did. We would then be left with quite a different conception of the natural kind gold.

The fundamental point to grasp here is that it is not a case of one of the sets of scientists being right and the other set wrong. How could we possibly say that one is right and the other wrong? The answer does not appear to be forthcoming from nature, as Putnam obviously thought it would be. It is the very ambiguity in nature which allowed for the setting up of the thought experiment in the first place. Both Earthian and Twin Earthian scientists must be judged to be equally correct. This is because nature is able to support both interpretations of the natural kind gold. Both isotope and atomic number can be judged to be equally important or fundamental properties. Taking either one as the basis for division will result in kinds or groups whose members share important regularities or similarities (similarity of chemical reaction with atomic number; similarities in radioactivity and stability/breakdown with isotope number). Each system seems to result in groupings which are equally natural and objective, in the sense that they highlight regularities which really are 'out there' in the natural world.

Donnellan's thought experiment therefore seems to lend support to my argument that categorisation is more than a reflection of the way the world is. The world appears to provide us with more than one way in which to categorise its entities/substances on the basis of important properties which are discoverable by science. It seems that, contra Putnam and other objectivists, the world is not uniquely structured and so does not fall into certain preordained natural kinds. Rather, it seems that the world is ambiguous – it can be divided up into kinds in different, yet equally natural, objective and scientific ways.

Human Input for Categorisation.

If it is the case that the natural world is, in itself, ambiguous and that it is able to support a number of different groupings of the entities/substances within it, then the kind of rigid realism espoused by Putnam and others is in trouble. The reason for this is that classification or grouping into natural kinds cannot simply be a matter of scientific investigation isolating *the* important 'natural kind' properties and then dividing the world into natural kinds according to these, thereby reflecting the one way in which the world is.

Rather, what Donnellan's thought experiment illustrates is that there can be more than one suitable candidate for the relevant 'natural kind' property, there can be more than one scientifically discoverable property which appears important or fundamental enough that it should be used to determine the nature and membership of a particular natural kind. If this is so, then scientists must be involved in a *choice* or *decision* as to which property should be taken as definitive of the kind. Our scientists do not simply reflect the hidden structure of the natural world, they actually play an active part in determining the nature of natural kinds.

In the Earth/Twin Earth example, Earthian scientists have, at some point in the past, made the decision that what is to count as *the* important physical property in delimiting the chemical elements is atomic number. Twin Earthian scientists, on the other hand, have decided that isotope number is what is relevant in order to divide chemical substances into kinds. Presumably, there will be a number of factors which might prompt the choice of one property over another. Donnellan points out that in reality, isotopes were discovered after the acceptance of atomic theory and that this timescale may have influenced the way in which we mapped our vernacular kind terms onto science. Timing of scientific discovery may well be one of the factors which influences what properties are taken as the important ones for the delimitation of natural kinds. However, Donnellan's point is that just because one important property is discovered before another does not mean that the first represents the one and only correct principle according to which the world can be divided. Time of discovery is a mere contingency and this is why it is possible to construct a Twin-Earth type story. We might imagine that chemical substances were originally divided according to atomic number on Twin Earth, but when isotopes were discovered, scientists considered that division according to isotope number constituted a more precise or accurate system which was better suited to their purposes; or perhaps that

isotope number was 'discovered' on Twin Earth prior to atomic number; or perhaps that both were discovered at roughly the same time, but that for some reason, isotope number was adopted in preference to atomic number. The motivations really do not matter. What is important is that there are different properties which could be taken as definitive of natural kinds. This indicates that our division of the world into kinds is not simply a reflection of the one way in which the world is and so ensures that human beings play some kind of active role in the classificatory process.

The Blurring of the Line between Metaphysics and Epistemology.

If we accept that Donnellan's thought experiment shows that there is more than one way in which we can divide the world into natural kinds and that it entails that human beings play an active role in the construction of natural kinds, then we have to admit that there is a blurring of the line between metaphysics and epistemology, between the way the world is and the way humans construe that world or have knowledge of it.

We have seen that there is more than one way in which the natural world is. It does not appear to have one inherent structure which dictates precisely how it should be divided or the kinds into which it should fall. Therefore, classification or division into natural kinds cannot *simply* be a matter of metaphysics. Of course, in a trivial sense, division into natural kinds is about the way the world is – it involves the grouping of entities in the world according to certain regularities or similarities which those entities share. However, since there is more than one way in which these entities can be grouped and since the world lends support to different divisions of itself, epistemology will also play a substantial role in classification in the sense that human beings must choose between these different classification possibilities, they must make a decision as to what is to count as relevant for the construction of natural kinds.

This kind of approach involves the assumption that the sharp distinction between metaphysics and epistemology which is embraced by the doctrine of Objectivism (as discussed in Chapters Two and Five) is misguided. It is simply not the case that the world comes with one preferred description and that it is the job of science to discover the underlying structure of the world and then passively reflect that structure. Rather, we need to acknowledge the complexity and richness of the natural world and realise that as a result, it will not be susceptible of only a single

description. Furthermore, we need to see ourselves firmly located within the world, as a part of it. Our actions, choices and decisions can then be construed as contributing to the way the world is, to its very fabric. Seen in this way, the division between metaphysics and epistemology must be misguided and human beings will inevitably play an active role in the definition and delimitation of natural kinds.

Of course, none of this is to say that it is open to us to divide up the natural world or to form natural kinds in whatever way we like. What we do will always be constrained by the world. That is, different entities or substances may share an infinite number of properties in common. However, only some of these properties will fulfil Putnam's description by being important physical properties of the type with which scientists concern themselves – usually microscopic and so 'hidden'. It is only these kinds of property which will be suitable or relevant for use in scientific definitions and descriptions. Furthermore, only some of these properties will result in any useful groupings for us. Members of the group or kind must share important characteristics; they must all exhibit certain regularities or similarities of behaviour which can explained by possession of the relevant scientific property. What Donnellan's story shows is that even given these kinds of stipulation, there is still room for alternative groupings, for alternative definitions of natural kinds. Twin Earthians have chosen isotope number to delimit the natural kind gold and this involves important regularities or similarities shared by isotopes of the same kind; Earthians have chosen atomic number and this too involves important similarities between substances with the same atomic number. Neither system can be shown to be wrong, since both are based on scientific properties and facts, on states of affairs which are objective, which really are exhibited by nature. What we have is an illustration of the richness and potentiality of the natural world, together with the fact that human beings (and so epistemology) play a role in dictating the nature and boundaries of natural kinds.

John Canfield.

John Canfield in his "Discovering Essence" aims to present a rival theory to the realism of Putnam and Kripke. This rival theory rests on the idea that essence is not *de re*, but *de dicto* and is couched in terms of a Wittgensteinian criterial account whereby essences are determined, not by reality, but by criterial rules of language. For example, on a criterial account something which was not H_2O could still be

called 'water' if it possessed the criteria by which we commonly isolate water – if it looked, tasted and felt like water. This would, however, be unacceptable on a realist account since an appeal to such criteria is irrelevant – what determines whether something is water for the realist is whether it possesses the hidden essential property of being H_2O.

Firstly, Canfield introduces a distinction between what he calls intentional and extensional discoveries. Intentional discoveries are 'discoveries that...'. So, for example, we might discover that the essence of gold is atomic number 79 (intentional discovery). All other discoveries are extensional. Thus, rather than discovering that it is the essence of gold to have atomic number 79, we might simply uncover a number of important properties which gold possesses, including its atomic number (extensional discovery). Canfield then asks the question, "But does the scientist merely discover that gold has the atomic number, thereby discovering essence; or does he, alternatively, discover *that* it is the essence of gold to have this number?"[229] He thinks that Kripke is arguing for the second intentional alternative – that by discovering that gold has atomic number 79, scientists in fact discovered that it was the essence of gold to possess atomic number 79. Canfield cannot, however, accept this interpretation. He does not think you can discover the essence of something *per se* – how would you know that a particular property or properties were the essential ones? As he points out, we have discovered all sorts of scientific properties of gold – its lattice constant, its melting and boiling points, its coefficient of linear expansion, its tensile strength, its Brinall hardness, that it is a face-centered cubic metal – yet it is unreasonable to claim that all of these are essential. He therefore concludes that some kind of *theory of essence must be presupposed* before one can move from the claim that gold has atomic number 79 to the claim that it is the essence of gold to be an element with atomic number 79. It is this kind of theory which enables one to isolate the essential from the accidental properties.[230] He is therefore opting for the extensional account.

Canfield observes that Kripke sometimes makes use of a particular strategy in defending his own account of natural kind terms. This involves constructing abnormal states of affairs and then asking, for instance, whether these particular

[229] Canfield, 1983, p.107.

[230] Note the uncanny similarity here with the psychological explanation-based model of categorisation, which claims that some kind of underlying theory or explanation is needed to account for why certain properties are deemed the important ones over and above others for matters of classification.

objects are tigers or whether this particular substance is gold.[231] Kripke claims that the answers given to these kinds of question go against criterial accounts of natural kinds. The idea is that we can imagine a case where criterial features are absent and yet we would still say that the entity is of the kind in question, or alternatively we can imagine a case where the criterial features are present and yet deny that the entity belongs to the kind in question. Criterial features cannot therefore determine kind membership. Kripke claims that these anti-criterial answers result from our intuitions.

Canfield, however, is unconvinced by this. He maintains that we can construct similar cases where our intuitions give results which *conform* to the criterial view and go against Kripke's own theory. He asks us to imagine a case where some animals which look very like what we call tigers are dubbed 'tigers' by explorers. Everyone else who subsequently comes into contact these animals sees them as possessing tiger-like characteristics. The idea here is that the word 'tiger' is *introduced* by these explorers – these people did not possess the word prior to this. In fact, there is an illusion surrounding these animals and really, under veridical conditions, they are like what we call chickens. What happens if the illusion clears? Canfield claims that we *might* say that there were in fact no tigers in this case but we might equally well say that there are tigers in this case, but they turned out to be somewhat different from what was originally supposed. Neither statement would be better or worse or truer to the facts than the other. Canfield stresses that he perfectly understands both responses and, furthermore, grants the truth of both. (Kripke would of course say that the only correct statement is the second, since he believes that criterial or surface features are irrelevant for determining kind membership.)

The reason why Canfield wants to say that either answer is equally correct is that, contra Kripke, he does not want to claim that this is a matter over which one can be objectively right or wrong. Rather, any intuitions which we have on the matter will arise from certain presuppositions involving the choice of a particular criterion for the extension of the normal use of kind terms to these abnormal situations – but there will also be alternative criteria which it would have been equally correct or legitimate to presuppose. If we claim that there are no tigers, then we are implying that something is a tiger only if it has at least some of a collection of specified features. Since the jungle tigers were an illusion, they did not in fact possess any of the

[231] See Chapter Two for some examples of this kind of strategy.

features originally attributed to them. If we claim that there are tigers, then we are implying that something is a tiger if it is the entity pointed to in the original reference-fixing act or if it lies behind our original misperceptions. Each criterion is legitimate and what intuitions we have will be determined by which criterion we accept. Kripke, however, would claim that only the second criterion is correct, since he has already made the decision to equate essence, not with surface features, but with hidden features of the (majority of) entities in the original sample.

Canfield goes on to extend this kind of argumentation to the example of gold. He imagines scientists examining the original set of samples called 'gold' in a particular society and discovering that most of them have atomic number 79. However, a few of the samples are of some substance which looks and behaves almost exactly like gold, except that it does not have atomic number 79. In other words, both substances have the same surface features but do not share the same 'hidden' features – the two substances are much more similar than real gold and iron pyrites or fools' gold are. For the society in question, this substance and the substance with atomic number 79 are interchangeable – for every practical purpose they are the same. In this kind of situation, Kripke would of course say that scientists have discovered that it is the essence of gold to have atomic number 79. Those samples in the original set which lacked this property are deviant and so are not really gold – the society was mistaken in labelling them 'gold'. Canfield, however, argues that it would be plausible to say either that the 'deviant' samples are gold or that they are not gold. It is as plausible to say that the two sorts of sample form a natural kind as to say that they do not. As he puts it, "One could say either, depending on what criterion for "gold" one presupposed".[232][233]

The problem with Kripke's theory, according to Canfield, is that he is presupposing what he wants to prove. This is because, "Everything depends on what we mean by, or count as, "the same kind of stuff"."[234] By claiming that two samples will only count as gold if they both have atomic number 79, Kripke has already presupposed

[232] Canfield, 1983, p.118,

[233] Zemach (1976) makes a very similar point. Contra Putnam, he argues that we might say that Earthian water (with chemical composition H_2O) and Twin Earthian water (with chemical composition XYZ) are both equally water and that there are simply two types of water in existence with different chemical compositions. He says, "Chemical constitution is not always decisive in determining our usage of substance names." (P.119.) In other words, it is plausible to say that H_2O and XYZ are both water and so form a natural kind, but it is also plausible to claim that only one of these is really water and that they therefore do not form a natural kind.

[234] Canfield, 1983, p.118.

that having atomic number 79 is what is to count in determining that two samples are both of the same stuff – gold. Kripke is jumping the crucial step of criterion choice and instead just presupposing a particular criterion and taking that to be unquestionably correct. As Canfield states, "The natural use of "natural kind" does not carve up the conceptual territory in the way that Kripke wants it carved up; and if we want to use "natural kind" to carve it up that way we shall have to put a stipulation on its use. We will be stipulating that two things are of the same natural kind only if they have the same *k*-properties."[235]

Canfield finally gives a positive account of how the criterial view of kind terms might account for the scientific discovery of essence. He bases his account on the scientific discovery and definition of viruses, taken from Sally Smith Hughes' book, *The Virus*. Following the discovery of viruses, Hughes tells us that the problem for scientists, "...was to discover the intrinsic properties of viruses rather than to characterise them in terms of technique determined ones".[236] Canfield assumes that for "intrinsic properties", we can read essential properties. It would therefore seem that we have a situation which closely fits the realist Kripkean theory, whereby viruses were originally isolated on the basis of extrinsic or surface features and then it later became the job of scientists to discover the underlying essence of these entities. Hughes continues:

> "By the 1950s, structural studies of viruses with the electron microscope and information about their nucleic acid content provided a meaningful basis for distinguishing viruses from all other types of infectious agents. From this time on, 'virus' was used with a meaning roughly comparable to that given it today. Modern definitions abound but most characterize the virus as an infectious, but not necessarily pathogenic, entity which is usually submicroscopic, which contains a core of either DNA or RNA covered by a protein or lipoprotein capsid and which reproduces exclusively within living cells."[237]

Canfield argues that this statement can be seen to fit the criterial account just as well as the Kripkean one. The idea is that scientists, on examining viruses, discover a number of properties which viruses were not previously known to possess. Some of

[235] Canfield, 1983, p.119.
[236] Hughes, 1977, p.113.
[237] Hughes, 1977, pp.113-4.

these properties will be used in constructing a definition or definitions of a virus. It is *only when a definition is provided and accepted* that the essence of a virus emerges, not before. So, scientists do not discover that *x* is the essence of viruses, but they stipulate that a certain discovered property or properties are to be *counted as* the essence of viruses. What it is important to grasp is that there are different possible definitions of 'virus' which might have been given. Canfield notes that the quotation from Hughes emphasises that there are a number of different extant definitions of virus. So, when the term 'virus' is used with a particular definition in mind, then by dint of that very definition, the essence of what it is to be a virus *in that context* will be invoked. However, if the term were used with a different definition in mind, then the essence associated with that alternative definition would be invoked.

The criterial account differs from the Kripkean one in that it includes an extra step between the scientists' discovery of hidden properties and ascription of essence. This extra step rests upon the extensional discovery of essence – only when a definition is constructed can essence be said to be discovered – we do not discover essence simply by discovering hidden properties. In other words, a discovered property will only become essential because it is defined as such.

Furthermore, on the criterial view and contra the Kripkean view, definitions other than the actual one are possible given the hidden properties discovered by science. Reverting to the example of viruses, Canfield argues that it would be possible for scientists to discover that all known viruses, including the ones in the original set, have an inner core of either RNA or DNA, yet not include this as a necessary condition in the definition of a virus. The point is that if scientists decide to include having a certain size and inner molecular structure as part of the essence of viruses, then they will, by that very decision, ensure that in the future they will never discover a *virus* which fails to exhibit those properties, although they may discover an entity of *similar* size and molecular structure which displays similar behaviour. Only if scientists decide to change their definition of 'virus' might this entity be considered a virus. However, if scientists had originally decided to designate different properties as essential to viruses, then this entity might straight away have been considered a virus. Canfield comments, "This illustrates the element of decision involved when a term is defined by means of properties science has discovered."[238]

Canfield's arguments for the criterial view are summed up thus:

[238] Canfield, 1983, p.123.

"Scientists may discover, say, that gold is an element and has the atomic number 79. It is then possible to define gold as the element with that number. When that definition is adopted, then it is of the essence of gold (on the use of "gold" so governed) to be of that atomic number, but whether to adopt that definition was a matter for decision, not discovery."[239]

Analysis.

Once again, we have a series of philosophical arguments which support my general thesis concerning categorisation and which lend theoretical weight to the evidence adduced from case studies in the previous chapter. Canfield's arguments and examples can be analysed in much the same way as those of Donnellan.

Categorisation as More than the Way the World Is.

One of Canfield's primary motivations seems to be to make the point that categorisation definitely comprises more than a simple reflection of the one way in which the world is. His contention is that no matter how much we know about the hidden structure of the world and its properties, simply knowing what this structure and properties are will not tell us what constitutes an essential property nor what constitutes a natural kind. This is because the world does not present us with exclusive candidates for essential properties nor unique groupings of the entities within it.

Why is it that the world cannot provide us with the unique answers which would support a theory of the Putnam/Kripke type? I have argued in previous chapters that the world is extremely rich and complex and that for this reason, there are different patterns of similarity or regularity running through it. Different patterns can therefore be made the basis of different yet equally valid, real or objective categorisations. Canfield argues along similar, if less explicit lines. Take his example of gold – the Kripke/Putnam position holds that it is the essence of gold to be an element with atomic number 79 and that this grouping and essence is a reflection of nature, quite independent of human decision and choice. Canfield points out, however, that although it is now accepted that to be gold is to possess atomic number 79, there are a number of other properties (see his list on my page 150) which could have qualified

[239] Canfield, 1983, pp.128-9.

as suitable candidates for the essential property of gold. These properties could equally have been contenders since they are all suitably scientific. That is, they appear to be scientifically relevant properties which capture important aspects or characteristics of the substance which we call 'gold'. If all these properties are roughly equivalent in some scientific sense, then we are in a position to claim that essence is not something which is determined solely by nature, since nature proves ambiguous concerning what is to count as *the* essential property; it is too rich or complex for one property to stand out. Rather, it seems that we are left with a *choice* between equally plausible and scientific properties. It is this choice which Canfield claims Kripke leaves out of his account – it is not a case of discovering that in actual fact or that in nature the essence of gold is atomic number 79, rather it is a case of discovering a number of important scientific properties possessed by the substance which we call 'gold' and then coming to a decision that one of these properties (or a number of them) is/are to count as the essential one(s). Categorisation or division into natural kinds therefore involves more than simply reflecting what nature dictates.

I assume that the point of the above example is that given the totality of the substance which we now call 'gold', there are a number of properties other than atomic number which we might designate essential without altering that totality. In other words, given a fixed group or class, nature is ambiguous regarding the essence of that group or class. However, Canfield's later arguments show that the boundaries of the group or class itself can also be ambiguous. We can see this in his thought experiment concerning the Kripkean original sample of gold where he suggests that the substance with atomic number 79 and that without it could equally be said to form a natural kind as not. The point here is that what property or properties we judge to be essential will dictate what belongs within and what is excluded from the group or kind. In Canfield's thought experiment, if atomic number 79 is judged to be essential, then the substance which does not possess this atomic number will not fall within the kind of gold. However, if we take some property which is shared by both substances as essential, then it is possible to say that both substances form a natural kind. Again, nature faces us with an ambiguous situation, replete with information, where the choice is not ready made for us. At some point we have to cut through this information and make a decision concerning what is to count as essential and it is this decision which will fix the boundaries of the natural world.

Looking at Canfield's discussion of the scientific discovery and definition of viruses we once again see that nature by itself is ambiguous and that dividing and defining of the natural world involves an element of human choice or decision – that categorisation involves more than just reflecting the way the world is. Since the world is so complex, we must select a particular aspect or facet of it and make that the basis of division. Canfield indicates that with viruses, as with other 'scientific' entities, scientists have discovered many 'hidden' or non-surface features which viruses possess. Which of these hidden properties is to count as essential? Canfield claims that this is something which must be determined by scientists, since nature provides us with a number of possibilities – we must decide between these. It is only once the scientists have made the decision that we are in a position to say what the essence of a virus is, since it is by dint of that very decision that the relevant essence comes into existence. Furthermore, which properties we take to be essential will determine which entities fall within and which fall outwith the group. As Canfield shows with the example of viruses, two entities which appear to be very similar may both be considered viruses, or one may be considered a virus and the other not, dependent on what definition of 'virus' has been adopted.[240] Canfield seems to be saying that it makes no sense to talk of preordained essences and kinds existing in the world, quite independent of human beings, since it is only when we make a decision to elevate a particular property or properties to the essential level that we can talk of essence existing. Although we discover hidden or scientific properties, when it comes to the stage of judging which are essential and which accidental, we are in the realm of decision-making and not discovery. Canfield even talks about essence being defined into being, again suggesting that essence does not exist *per se*, but that we manufacture what is to count as essential through our choice of certain definitions and properties over and above others. He says, "Certain things called "viruses" are examined scientifically and found to have a number of properties, previously unknown. Some of the properties are utilised in constructing one or more definitions of what a virus is. At the point where a definition is given and accepted, essence is established. Essence is defined into being."[241]

Again, I think it is extremely important to stress that Canfield does not want to claim that we create the world in which we live, that all knowledge is subjective or anything along those kind of lines. His account is, in fact, firmly rooted in reality and

[240] This, of course, is a parallel to the thought experiment about gold discussed in the paragraph above.
[241] Canfield, 1983, p.120.

he himself comments that of course he is not saying that we can adopt any definition of 'virus' that we like. The point is that scientists do discover all sorts of scientific properties about viruses, gold, or whatever – properties which were not previously known about. There is no doubt that what we call gold does have atomic number 79 or that what scientists call viruses do have inner cores of RNA or DNA. These are real properties which are objectively there, and can be discovered given sufficiently advanced science and technology – they are not created by us. Therefore it does make sense to say that science discovers these properties. However, where Canfield wants to argue with philosophers such as Putnam and Kripke is over the idea that we discover the essence of something in the same way that we discover its properties. The claim is that we cannot simply discover that *x* is the essence of *Y* because there are several properties which might be counted as the essential ones – nature does not come with classes and essences affixed. So, it is at the level of formulating definitions, deciding on essences and so drawing boundaries that the human contribution comes into play. Of course hidden and scientific properties will restrict the definitions which we come up with since it is on one or some of these very properties that scientists base their definitions. Canfield notes that Hughes' use of the words "accurately defined" still make sense on his criterial account – once the hidden properties of of the group of entities which we call 'viruses' have been discovered by science, any definition of a virus which is inconsistent with those properties will count as inaccurate. In order to be accurate, a definition must comply with those properties. Yet there is still a gap here – discovering these properties does not mean that we discover that certain of them are already labelled as essential in nature. Rather, we must make our own choice as to which are essential and it is this choice which will fix the boundaries of natural kinds.

So, Canfield's arguments seem to fit neatly with my own contention that the natural world has different patterns of similarity running through it, even at the scientific level and that this necessitates our choosing one of those patterns and using it as the basis for our division of the natural world. This account remains a realist one, since it is grounded in properties which are real and objective and yet it imports an active role for human beings, since the world is not seen as coming with a ready made description as philosophers such as Kripke and Putnam would have it. Canfield illustrates that conflicting definitions and hence conflicting divisions into classes can be simultaneously correct in the sense that they can all be equally objective, equally based on scientific properties and so equally well-supported and justified.

Human Input for Categorisation.

If, as Canfield says, we do not discover essence, rather we decide what is to count as essence, then categorisation cannot purely be a matter of metaphysics or reflecting the way the world is. Rather, the classes and essences into which we divide the natural world must involve a certain amount of human input in the form of making choices and decisions.

In fact, one of Canfield's main complaints against a Kripkean type account of essence and natural kinds is that it fails to give any credit to this human decision making element. Canfield inserts an extra step between the discovery of certain hidden properties by science and the 'discovery' of essence and it is this extra step which introduces the notion of human decisions and choices. Whilst Kripke claims that in discovering the hidden properties of *Y*, scientists by that very discovery also discover *Y*'s essence (intentional discovery of essence in Canfield's terminology), Canfield claims that scientists discover *Y*'s hidden properties but that they then make a decision as to which property or properties will figure in definitions of *Y* and by that decision they elevate that property or properties to the essential level (extensional discovery of essence). The point is that given the set of discovered properties, other definitions could reasonably have been given which deemed alternative properties in the set essential. Given this kind of ambiguity, Canfield is saying that some kind of human input is necessary to make sense of any theory of essence.

It seems to me that the reason Kripkean and similar accounts do not make any mention of human input for categorisation is that such accounts always concentrate on the definition which is *currently accepted* and then work backwards from there.[242] This sort of approach will, of course, yield the impression that essences and definitions are fixed unambiguously by nature, since the accepted definition (if there is one) is always assumed to be the one and only correct definition or it is assumed that there is one such definition but it has yet to be discovered. This is why it is so useful to look at case studies, as I have done in the previous chapter, since these provide us with information which goes beyond accepted definitions and so make it easier to see how the kind of account mooted by Canfield actually fits the process of definition which takes place in science.

[242] The examples used to back up such accounts are also carefully chosen to be ones where (a) the definition is well-established and (b) where there is only one extant definition and so no question of multiple or conflicting definitions.

Canfield himself makes much the same criticism of Kripke. He claims that all Kripke is in fact doing is choosing one particular criterion for the definition of a natural kind (the one currently accepted in science) from a number of other possible criteria and then claiming that this is simply a matter of discovering the essence of that natural kind as dictated by nature. For example, he assumes that atomic number 79 is the discovered essence of gold but, according to Canfield, he is in fact choosing atomic number 79 as the criterion for deciding what is and what is not gold (or is following the choice which has already been made by others). Kripke is already imposing some kind of stipulation on what is to count as the same substance – gold, in this case – but it is important to realise that this stipulation is man-made and not nature-made. Given the atomic number 79 stipulation, Canfield's two indistinguishable substances with different atomic numbers will not be allowed to count as the same substance, yet given an alternative stipulation, it may be acceptable to count them as the same substance. What it is important to grasp here is that either decision is potentially correct but what makes the one acceptable and the other not is the human stipulation or decision as to what is to count as being the same substance. On this account, Kripke's theory simply becomes an illustration of Canfield's criterial account.

Likewise, the kinds of abnormal cases which Canfield quotes Kripke as constructing in support of his theory are simply the products of stipulations concerning what is to count as the same. Kripke claims our decisions in these kinds of situation reflect our intuitions and that our intuitions therefore conform with his theory. Canfield, however, is of the opinion that our 'intuitions' simply reflect an extension to the abnormal of what would be said in the normal case and so reflect the stipulations which are attached to the normal case. So, for example, if all cats are defined by scientists as possessing a certain genetic code, then we might say that something which possessed that code but looked like a dog was still a cat and not a dog. But this decision would only be a reflection of the stipulation that what makes a cat is possession of the relevant genetic code and so could not be used to support a Kripkean type theory of natural kinds. Alternatively, if all cats are defined on the basis of surface features and not genetic code, then we might say that the entity which looks like a dog can, under no circumstances, be a cat. But this is just reflection of an alternative stipulation governing the use of the term. We might say that different properties are made salient, dependent on what stipulation is placed on the use of the term. Yet what we take to be salient and so what that stipulation is is not a function of the one way in which the world is, since the world is such that it

admits of a number of different divisions and definitions. Rather, this is a function of human decision as to which of many properties is/are to count as essential in relation to a particular kind or class. It is this human element which is excluded from accounts such as Kripke's, yet it is this element which is fundamental given the richness and consequent ambiguity of the natural world.

The Blurring of the Line between Metaphysics and Epistemology.

Much of what needs to be said in this section has already been stated in the analysis of Keith Donnellan's paper above. As before, the basic point is that if human beings play an active role in the categorisation process by making decisions and choices, then categorisation is not simply a matter of discovering and reflecting the inherent divisions and essences of the natural world – categorisation is not simply a matter of metaphysics or the way the world is. Rather, it involves a *combination of metaphysics and epistemology*. That is, on the basis of metaphysical properties which are discovered by science, humans make decisions or choices as to which of those properties are to figure in definitions or which are essential and this in turn specifies a rule which can be used to determine which entities fall within and which fall outside the class. And, running alongside this, is the notion that alternative choices and decisions could have been made.

Canfield's account can clearly be seen as supporting the argument that categorisation comprises a mixture of metaphysics and epistemology. He acknowledges that scientists discover all sorts of 'hidden' or 'scientific' properties of entities and substances and that these are objective and real properties of those entities and substances – this is where the metaphysics comes in. What he denies, however, is that scientists thereby discover the essence of these things – he does not think that they discover property *x* and thereby discover that property *x* is the essence of *Y*. In fact, he argues that essence is not discovered in this way at all. Rather, essence is a matter of decision, in the sense that after having discovered all sorts of properties of a given entity or substance, scientists will then go on to raise one or some of these properties to the essential level by incorporating them in a definition. Which properties the scientists nominate essential will determine what qualifies as the same sort of entity or the same kind of substance. This is where the epistemology or element of human input shows itself. Since the natural world is ambiguous, this human input is necessary in order to get any account of categorisation, natural kinds or essentialism off the ground.

Conclusions.

As we have seen, Donnellan's and Canfield's accounts tie in with and provide support for my own arguments as expressed in the previous chapter. In particular, they provide arguments for and illustrations of the idea that natural kinds and their essences are not solely determined by nature itself. Rather, nature presents us with a number of ways in which we might proceed rather than mapping out one single path for us to follow. As a result of this, natural kinds and their essences must involve a degree of human choice and decision, something which neither Putnam nor Kripke wishes to allow.

We have therefore now reached a fairly strong position. We have evidence from a number of scientific case studies which illustrates the contention that categorisation comprises a mixture of metaphysics and epistemology. We also have philosophical arguments which show why categorisation cannot purely be a matter of metaphysics and which thus serve to underpin the evidence gleaned from case studies.

Chapter 8
Expert Scientists and Explanation.

Introduction.

The purpose of this chapter is to further clarify some of the issues which have been discussed in this thesis.

I aim to give a more theoretical grounding for the claim that the explanation-based view applies as much to expert scientific as to lay classification. This provides a philosophical basis for the supporting evidence gleaned from the case studies in Chapter Six. I also aim to show how I was initially led to apply this model to scientific classification. In order to do this, I argue that Putnam's hypothesis of the division of linguistic labour is a satisfactory account of people's behaviour and show that his hypothesis has been taken on board by psychologists of categorisation. Despite this acknowledgement of the role which scientists play, psychologists have not analysed the ways in which scientists classify nor the form which their concepts take. One of the goals of this thesis has therefore been to redress the balance and to show that the explanation-based account is a suitable model of scientific classification.

Secondly, I examine the ways in which the scientist is constrained in his categorisation of entities in the natural world. I argue that two very important constraints are that of utility and that of what counts as a 'suitably scientific' property and that these two constraints are linked by the mechanism of explanation. Finally, I explore the nature of the connection between explanation and classification.

Division of Linguistic Labour and the Explanation-Based Account.

Putnam's Hypothesis.

> "Every linguistic community...possesses at least some terms whose associated 'criteria' are known only to a subset of the speakers who acquire the terms, and whose use by the other speakers depends upon a

> structured cooperation between them and the speakers in the relevant subsets."[243]

The above quotation encapsulates Putnam's division of linguistic labour. The general idea behind this hypothesis is that there are a (large) number of terms in our language (the number has increased with the advent and growth of science) which are used quite successfully and correctly by non-expert laypeople. However, the correct use of these terms does not require the speaker to be familiar with the (expertly/technically) correct definition or meaning of the term nor to be able to recognise or distinguish those entities or objects to which the term refers from other (perhaps superficially similar) entities or objects. The reason for this is that the linguistic labour is divided between experts and laypeople. The definition of certain terms will be known by the relevant experts, who will also have the requisite knowledge and skills to isolate and recognise the entities or objects to which the terms refer. Expert knowledge thus provides a back-up for the layperson's use of a term and in the event of a dispute over the correct use of a term, (whether *x* really is an *F)* the relevant expert, possessing the requisite knowledge, can be relied upon to solve the dispute.

Putnam illustrates this example with the use of natural kind terms. We are all able to use terms like 'gold' and 'water' and have a good knowledge of the surface characteristics of both substances which, under everyday circumstances, allows us to state that our wedding ring is made of gold or that there is a glass of water on the table. However, if we had to distinguish 'true' gold or 'true' water from superficially similar substances, we would be in trouble. The expert scientist, however, (probably the chemist in this case) would be able to solve the problem, since he is acquainted with the relevant 'scientific' properties which make something 'really' water or 'really' gold.

Putnam concludes:

> "Whenever a term is subject to the division of linguistic labor, the 'average' speaker who acquires it does not acquire anything that fixes its extension...it is only the sociolinguistic state of the collective linguistic body to which the speaker belongs that fixes the extension."[244]

I think there is little doubt that Putnam's hypothesis is correct. We certainly do rely

[243] Putnam, 1975, p.228.
[244] Putnam, 1975, p.229.

on the knowledge of experts and the definitions with which they can provide us. It seems that part of our concept of gold, for example, does comprise the definition that the relevant experts attach to that term. And, in everyday life, we can continually see Putnam's hypothesis in action. If we want to know what species an animal belongs to, we might consult a biologist or a vet who has the requisite biological training. If we want to know whether a plant belongs to one species or another, we might ask a botanist. If we dug up something with the appearance of a diamond or a piece of gold, we might go to a chemist to ascertain whether it really was diamond or gold.

I do not think that Putnam's hypothesis conflicts with the evidence and arguments which I have adduced to show that scientific classification comprises more than a reflection of metaphysics and that there is more than one way in which the natural world is. None of this alters the fact that scientists know a lot more about the hidden microstructure and scientific nature of entities than does the average layperson. In view of this and in view of the fact that there is a widespread acquiescence to the notion that such factors as hidden microstructure comprise the real nature of things, one would expect laypeople to rely on scientists to back up and confirm the judgements which they make. As noted at the beginning of Chapter Six, most people are simply unaware of the uncertainty and disputes which arise in the domain of scientific classification. However, this is quite a different issue from how people do, in actuality, behave and, in terms of people's *behaviour*, I believe that Putnam's hypothesis is shown to be correct.

The Explanation-Based Account.

When looking through the psychological literature, it can be seen that with the rise of the explanation-based view, there seems to have been an acknowledgement of the notion that expert opinion plays some sort of role in the layperson's conceptual representations of natural kinds. Perhaps the most explicit statement of this occurs in the explanation-based model known as psychological essentialism, which I discuss in Chapter Four of this thesis. To recap, Medin and Ortony claim that at the deepest level, laypeople's conceptual representations involve an essence placeholder which can be filled with, amongst other things, the belief that there are experts who are acquainted with the relevant essence, even if the layperson himself is not.[245] Lloyd Komatsu notes the connection between Putnam's hypothesis and the explanation-based view in his paper summarising recent theories of conceptual structure. He

[245] See p.79 of this thesis.

writes, "I argue that some (but not all) of the issues raised by Putnam's analysis distinguish the explanation-based view from the similarity-based views."[246] There is also some psychological experimental evidence which supports the notion that people defer to expert opinion regarding the nature or definitions of natural kinds. For example, Barbara C. Malt reports that during an experiment, subjects were told about a plant that was between a marigold and a dandelion in appearance. They judged that with respect to that plant it made more sense to say, "We'd have to ask an expert to tell us which it is" than, "I guess you can call it whichever you want".[247]

So, we have the situation where Putnam's hypothesis is being taken seriously by psychological researchers dealing with (specifically explanation-based models of) concepts. This insight has led me to move up a level and pose the question, *what form do the concepts of the experts take?* There is no real guidance from the psychological literature here, since, as I have noted before, most psychological work on categorisation and concepts concentrates on laypeople. There does, however, seem to be an implicit assumption coming at least from some psychologists that expert categorisation is different from lay categorisation. In particular, there seems to be a suggestion that expert scientists are acquainted with metaphysics or the way the world really is and this therefore seems to tally with the idea that the way in which they classify should fit with more traditional models of categorisation. This is a point which I discuss on pages 95-96 of this thesis with respect to the interchange between Rey on the one hand and Smith, Medin and Rips on the other. This point is further emphasised in a paper by Edward E. Smith in which he writes, "...a metaphysical categorization of an object with respect to *gold* would take atomic structure to be the ultimate criterion – because that is the best guess of the relevant science as to what the ultimate criterion is – whereas an epistemological categorization of an object with respect to *gold* might consider the object's color, value and common usage."[248] There is also, of course, the assumption made by laypeople that entities possess essences which make them the things that they are and that it is the job of the expert scientists to pinpoint that essence. This is illustrated in the paragraph above and by Rips' and Keil's experiments, which are discussed in Chapter Four of this thesis. The general picture, then, seems to be one of scientific experts who classify in a very different way from laypeople and whose concern is with metaphysics and not epistemology.

[246] Komatsu, 1992, p.501.
[247] See Malt, 1990 and Malt, 1994, p.42.
[248] Smith, 1989, p.57.

This picture, however, is at odds with the picture that I have been building up of the way in which scientists classify. My claim throughout this thesis has been that scientific classification does not fit an objectivist or classical type model and that it is not purely a matter of metaphysics.

What is wrong with traditional views of categorisation for experts?

We have seen throughout the course of this thesis that positions such as Putnam's early stance concerning categorisation and natural kinds involve a number of implicit but fundamental assumptions. In particular, I explained in Chapter Two how such accounts are underpinned by a doctrine which I labelled Objectivism which roughly states that the natural world possesses a unique inherent structure and that therefore there is only one 'correct' way of carving up this world – that is in accordance with this structure. This doctrine involved various cognate assumptions such as that classification is purely a matter of metaphysics and that human beings provide no active input for the classification process.

The remainder of this thesis has involved showing that these objectivist assumptions are misguided with respect to classification at what might be regarded as the most fundamental level – the level of the scientific expert. You would expect that if Objectivism were a correct interpretation of both the natural world and the human act of classifying the entities which make up that world, then this would be manifested at the scientific level by scientific experts whose aim it is to uncover the nature of the world. Indeed, both Putnam and Kripke in their objectivist accounts of natural kinds and essences align their theories with the findings of science. However, in the second part of Chapter Two, I discuss some general problems with the doctrine of Objectivism which centre on the notion that this doctrine makes no sense or is logically impossible. In Chapters Six and Seven, I discuss the assumptions of Objectivism specifically in relation to the scientific expert. Chapter Six looks at scientific case studies and the findings reported there show that the way in which scientists actually go about carving up the natural world is not consonant with Objectivism. In Chapter Seven, we saw that there are various philosophical arguments which can be deployed against the objectivist conception of natural kinds and essences. The most important point to retain from all these studies, I think, is that it is not possible to draw a clear line between metaphysics and epistemology in the way that the objectivist wants to do. Once you have shown that this is an impossibility, you have removed the linchpin of the doctrine and the whole structure

comes tumbling down.

Towards an Explanation-Based Account for Experts.

Paving the way.

There are two main points which, in my mind, lead towards the idea that the explanation-based account is a suitable model of categorisation by scientific experts. The first is the point made above – it is not possible to make a clear demarcation between metaphysics and epistemology nor to claim that there is only one 'correct' classification of the natural world which involves mirroring the world's inherent structure. This, as I have repeatedly argued before, is because epistemology or active human input is an integral component of classification. In fact, the epistemology or human input is an absolute prerequisite for getting classification off the ground.

Why is the epistemology so vital? This leads on to the second main point in support of the application of the explanation-based view to scientific experts. This is that the natural world is extremely rich and complex, which means that there are different patterns of similarity or regularity which cut across this world and which can be picked out and used as the basis for dividing the world into kinds or classes. It is at this juncture that epistemology becomes important – given these different patterns of similarity or regularity, it is necessary that the human agent make some kind of choice or decision as to which pattern is to be taken as salient and so which properties are to be taken as definitive for matters of classification. Without this degree of human input, classification remains an open or underdetermined question. These points came out particularly clearly in relation to the case studies. There we saw a number of situations where a decision needed to be and was made by a scientist or scientists regarding which patterns and properties were to count as relevant for or definitive of a particular class. In the case of Dean's study, we even had a situation where different schools of scientists advocated entirely different systems or sets of principles according to which the plant kingdom should be divided.

How can we apply an explanation-based view to experts?

It is helpful here, I think, to briefly recap on the structure of the explanation-based account. As a general psychological model it is currently somewhat underspecified. However, there are a number of general points which make up the body of this

account. These include:

- Entities are not similar *per se* – the respect(s) in which they are similar must be specified. This involves some kind of underlying theory or explanation which accounts for why or in what respects we regard certain entities as being similar.

- Concepts do not simply comprise lists of features or attributes and so categorisation is not simply a matter of attribute-matching. Again, some kind of underlying theory or explanation is involved which specifies why those particular features and attributes are taken as definitive of or relevant to the concept and which relates these features to one another.

- The explanation-based model is a *relational* account. As stated above, the features associated with a concept are underlain by some unifying theory or explanation which accounts for the salience of those particular features. Furthermore, concepts themselves are not isolated entities – much of their meaning comes from the ways in which they are related to and interact with other concepts.

Glancing over these points, it is immediately obvious that the explanation-based view is leaning away from an objectivist interpretation of the world. In particular, there is a marked move away from a conception of the world and our cognition of it as fixed and static. Rather, the emphasis is upon an inherent flexibility and dynamism, leaving room for human input in the form of explanation and theorising.

How, then, can the explanation-based view be applied to classification carried out by scientific experts? Well, the key lies in the points made in the foregoing section that, due to the complexity and richness of the natural world, scientists are forced to make a choice or decision regarding how and on the basis of which properties the natural world is to be divided into kinds or classes. When such a choice is made, there will be certain theories or explanations underlying that choice which account for particular properties being counted as relevant for similarity and categorisation judgements and for particular definitions being applied to particular kinds. This phenomenon was discussed in detail in Chapter Six where we saw from the case studies that the natural world is able to support competing classifications of the entities within it and that choice of a particular classification or classificatory system

involves and so can be explained by prior adherence to specific aims, theories and commitments (the explanations of the explanation-based account). It is these aims, theories and commitments which serve to make certain properties and similarities salient over and above others for matters of classification.

There is also an interesting parallel to be drawn between the explanation-based account of categorisation as I am applying it to scientific classification and John Canfield's account of essence as presented in Chapter Seven. To recap, the core of Canfield's argument is that discoveries of essence are extensional – scientists discover a number of properties of a kind and they then deem one or some of these to be essential to that kind. Since ascription of essence (as opposed to discovery of properties) involves some kind of decision on the part of scientists, Canfield is of the opinion that *a theory of essence must be presupposed* before any statements regarding essence can be made. Even the terminology[249] used by Canfield is reminiscent of explanation-based accounts. What both he and the explanation-based theorists are getting at is that it must be specified in what way or why entities count as similar. The natural world itself simply has too many similarities running through it or is too unconstrained to offer a unique answer. It is therefore up to us (or, more specifically, human scientists) to state what properties are to count as essential, or what properties are to count in terms of judgments of similarity for classification. It is then the reasoning behind this statement which comprises the theory of essence or the underlying explanation.

Summary.

What we have, then, is a situation where the explanation-based model can be seen to apply at the layperson's and the scientific expert's level of categorisation. We have seen (Putnam) that there exists a division of linguistic labour in society, whereby it is assumed that the relevant experts know the definitions attached to natural kind terms and can therefore be relied upon to solve disputes. There has also been a suggestion (from the psychology) that people tend to assume, particularly regarding natural kinds, that there are experts who will be able to define or specify the essence of the kind, even if these people are themselves unable to do so. Furthermore, I think there tends to be a general assumption that the world displays only one true nature and that it is the job of scientists to discover what that nature is. What I have argued during

[249] Some explanation-based theorists talk about 'theories' rather than 'explanations'. See for example, Murphy and Medin, 1985 and Keil, 1989.

the course of this thesis, however, is that the world is able to support different groupings of the entities within it and therefore that even at the level of expert classification, theories and explanations play a fundamental role in determining how entities are are to be classified and which properties are to be taken as salient. Even at the level of the scientific expert, classification is not purely a matter of reflecting metaphysics and epistemology plays a key role.

Limitations on Explanations.

It is important to notice, however, that despite the fact that the scientific expert must play an active role in carving up the natural world, this *does not* mean that that the scientist can carve up the world in any way that he likes, attributing any essence whatsoever to any kind which he chooses to construct. This is a point that has been stressed throughout this thesis in a general way. I think it is now useful to look at specific ways in which the scientist might be limited in dividing into kinds, attributing essences and utilising explanations. I suggest that there are (at least) two important ways in which the scientist will be constrained and that these are firmly linked together.

The Limitation of Utility.

The first is a point that has been made earlier in this thesis. This is that the scientist must be bound by *what is useful* in the ways in which he classifies and divides. What this means is that the job of the scientist is to uncover and exploit *regularities which really do occur in nature* – which really are out there and are not just products of our imagination. It is these regularities which must serve as the basis for the division of the natural world into classes. It is in this way that the scientist is bound by metaphysics or the way the world is in his classifying. The reason why all of this is bound up with utility is that to discover and exploit regularities in the natural world is to make that world something which we, as human beings, can negotiate or successfully interact with. To know that certain certain regularities hold between entities is to to be able to predict how those entities will behave. To know that a certain substance has atomic number 79 is to know that instances of that substance will behave comparably under certain circumstances. It enables us to foresee that behaviour and to act accordingly. On the other hand, to know that two instances of this same substance may have *different* isotope numbers (but the same atomic number) is to know that the two may behave in *very different ways* under a different

set of circumstances or conditions. Again, this enables us to predict and take account of a different kind of behaviour. In each case, we have some kind of regularity which occurs in nature. Possession of a particular property results in some kind of uniformity or predictability of behaviour and recognition of that fact enables us to negotiate or interact with that part of the natural world.

The mistake of the objectivist, then, is not in claiming that classification is concerned with metaphysics. Of course it will be concerned with metaphysics – it is the job of the scientist to uncover the regularities and patterns which *in reality exist in the natural world.* Rather, the objectivist's mistake is to assume that there is one unique set of regularities and so classes into which the natural world can be divided. The point that I am trying to make is that *even at the scientific level* there are different patterns and regularities which crisscross portions of the natural world. These can underlie differing divisions of the world, dependent upon which regularity you are concerned with or wish to make salient, as is illustrated in the atomic/isotope number example above. It is not that we, as human beings, are untidy in our cognition of the world, rather it is that the world is untidy – it does not neatly organise itself into one unique set of classes – it is simply too information-rich for that.

The 'Suitably Scientific' Limitation.

The second way in which I envisage scientists being constrained in their classification is that those properties which are deemed to be essential for or definitive of a kind *must be suitably scientific.* Although it is very difficult to give a clear definition of 'suitably scientific', what I roughly mean is that these properties must be at a theoretically fundamental enough level that they are suitable candidates to play a role in scientific explanations of important physical behaviour and characteristics. To illustrate: to group entities together on the basis of whether they are pretty or not is unlikely to qualify as a suitably scientific classification. This is so for a number of reasons. Firstly, entities which are all pretty are unlikely, as a group, to exhibit important regularity of behaviour or characteristics beyond their prettiness. In other words, such a grouping would be more-or-less arbitrary from the scientific point of view. What important or significant regularities can we find in the class of a pretty gemstone, a pretty flower and a pretty painting beyond the fact that they are all described as pretty? Secondly, prettiness is a non-measurable characteristic and so is unlikely to qualify as a good candidate for an essential property in a scientific explanation. In general, those properties which feature in the scientific classification

of entities are, at least to a certain extent, measurable or quantifiable in physical terms in a way that properties such as prettiness are not. It might be the case that a property such as prettiness can be accounted for by a scientifically more fundamental or significant property, but not vice versa. It may well be that the members of a scientifically relevant class are in fact all judged to be pretty, but this will need to be explained by the fact that the members share some other, more fundamental property or properties.

Different philosophers have used slightly different ways to describe the kind of properties with which scientists concern themselves. D.H. Mellor notes that advocates of essentialism tend to take microstructural properties as being the essential ones (atomic number for chemical elements, genetic makeup for plants and animals). He explains that this is because scientists use a principle of "microreduction" which involves the notion that, "properties of things should be explained in terms of the properties and relations of their spatial parts."[250] John Dupré defines a natural kind as, "a class of objects defined by common possession of some theoretically important property (generally, but not necessarily, microstructural)".[251] Putnam considers natural kinds to be groups of things which share "important physical properties".[252] There is an agreement, then, that a scientific classification will in some way be an important and relevant classification, in the sense that those properties which are deemed essential will be deeply explanatory of or fundamental to the class in question. Of course, what kind of properties are taken as relevant will vary between the sciences. So, microstructural properties play a very important role for classification in chemistry, whilst classifications in botany and zoology tend to make more use of properties which are at higher levels than the microstructural. Nonetheless, they are all likely to be suitably scientific candidates. In some sense, then, the 'suitably scientific' constraint upon the scientist is a psychological constraint. That is, it involves various people's[253] conceptions of what kind of properties the scientist ought to be concerning himself with. However, it is also a real or physical constraint. The scientist is aiming to discover the nature of the world and it seems that this nature is to be found at a deep or fundamental level of explanation.

[250] Mellor, 1977, p.310.
[251] Dupré, 1981, p.68.
[252] See, for instance, Putnam, 1975, p.239.
[253] Scientists themselves, philosophers and perhaps even laypeople.

The Role of Explanation.

To avoid confusion, it should be noted that in this chapter I use the word 'explanation' in two quite different senses. Above, I talked about explanation-based accounts of categorisation. The idea here is that there are underlying aims, theories and commitments which explain why properties are taken as salient and why categorisation judgements are made. It is the choices and decisions of the human categorising agents which are being explained. However, in this section, I talk about possession of essential properties explaining regularity of behaviour and characteristics. The idea here is that possession of these properties can be seen as the cause of things being or behaving in a particular way. It is the behaviour of entities in the natural world which is being explained.

How, then, are these two points – the point that scientists concentrate on patterns of regularities which exist in the natural world and the point that scientists deal in properties which are suitably scientific – linked? The key to this lies in the notion of *explanation*. We might say that the reason that entities have similar characteristics or behave in similar ways under similar conditions is that they share the same essential or fundamental properties. In other words, the possession of certain essential properties is used as an *explanation* of why these regularities occur. Possession of fundamental properties determines behaviour and characteristics.[254] This is again of course tied up with utility – it is useful in terms of interaction with and negotiation of the world for us to group together entities which react in similar ways, as explained by possession of common properties.

If we accept the key role that the notion of explanation plays, we then need to pose a couple of important questions. Does it make sense to claim that there is one ultimately basic set of scientific explanations of the natural world? And can that set be captured by providing a unique division of that world into kinds on the basis of possession of essential properties? I suspect that the objectivist will want to answer both these questions in the affirmative. He will want to say that there *is* one unique division into kinds and that this division reflects the most basic properties and so explanations of that world.

[254] This notion of explanation has an interesting parallel in Medin and Ortony's explanation-based model of psychological essentialism. They claim that people believe that surface properties of entities are constrained by deeper properties. This is why surface similarity generally works as a good heuristic for grouping entities and produces results which tend to correspond to the psychological essence. See pp.77-79 of this thesis.

This, however, is a move which I do not accept. My claim is that even at the scientific level, when classifications are made using important physical properties or microstructural properties or theoretically important properties, there is not necessarily a unique taxonomy of the natural world. I suggest that even at this level, the world still displays an inherent flexibility and indeterminacy. This means that there are alternative patterns or regularities (characteristics or behaviour) running through nature which are to be explained scientifically by different important physical properties. Since there are a variety of patterns and properties cutting across entities, it is possible to produce different, but equally objective, fundamental and scientific classifications of the environment.

Let's take a look at a couple of examples which reinforce this point. Keith Donnellan's example using gold which I talked about in the last chapter is a good example here. To recap, the basic thrust behind his argument was that although we group all isotopes of gold together as one natural kind on the basis of atomic number, it would be equally reasonable and scientific to divide what we call 'gold' into a number of natural kinds on the basis of different isotope numbers. As he points out, different isotopes of the same substance display important behavioural differences, particularly in relation to radioactivity and stability, whilst their behaviour may be very similar in relation to chemical reactions. Here we might say that the fact that all isotopes of what we call 'gold' possess atomic number 79 or have 79 protons in the nucleus of the atom is the important physical property which *explains* why they behave uniformly with respect to chemical reactions. However, it would also be correct to say that the fact that different isotopes of what we call 'gold' have different combined numbers of protons and neutrons in the nucleus of the atom *explains* why they behave differently regarding radioactivity and chemical breakdown. Each would seem to be a perfectly reasonable and correct explanation of the behaviour and characteristics of these substances. Both are scientific – they concentrate on microstructural properties and use these to explain what seem to be important characteristics and regularities of behaviour. How, then, can we say that one is more fundamental or basic or scientific an explanation than the other? It seems that we must admit that there are two different bits of scientific explanation going on. It would therefore be equally legitimate, objective and so correct to take *either one* as indicative of grouping into natural kinds. In opposition to the position of the objectivist, there is no call for claiming that one is fundamental and so that only one can be indicative of the uniquely correct natural kind.

A second illustration can be found in the later work of Hilary Putnam. This example is somewhat different from the one above since it is taken from the realm of biology and deals with complete methodologies for determining natural kinds and their essences. It also concentrates on two different explanations which can be given for the same set of characteristics and behaviour, rather than on two different explanations of two different types of behaviour, as in the example above. Nonetheless, the parallel is clear. Putnam is talking about two different conceptions of the essence of doghood – the evolutionary biologist's conception and the molecular biologist's conception. For the evolutionary biologist, species are historical entities and something will count as a separate species if it has achieved sufficient genetic isolation from other species. He can therefore be seen as studying populations and not individuals. As Putnam puts it, "Evolutionary theory seeks to understand the historical origins of the populations of animals that we see in our world, and to generalise about the mechanism of population change."[255] The molecular biologist, on the other hand, is not concerned with the history of populations. Rather, his concern lies with the *current* internal structure of organisms, with their RNA and DNA. Given these different concerns, Putnam then points out that the evolutionary and the molecular biologist take the essence of doghood to involve very different attributes or characteristics. For the evolutionary biologist, an explanation of what it is to be a dog will include the fact that dogs are descended from wolves and so, "being a domesticated wolf-or-hyena modified by millennia of selective breeding [will be] the essence of doghood".[256] This fact, however, is irrelevant to the molecular biologist's conception of the essence of doghood. His explanation of what it is to be a dog will involve the notion that dogs have a certain kind of DNA, a certain kind of genetic structure.

Putnam points out that these two different conceptions of what is to count as the essence of doghood will sometimes result in competing decisions about counterfactual situations. He devises a counterfactual case where we are invited to imagine that technology has become so advanced that it is possible to "synthesise a whole dog, starting from chemicals on a shelf" with exactly the same DNA as his own pet dog. The idea is that this new dog would be considered a dog by the molecular biologist but not by the evolutionary biologist. Putnam says, "I suspect, in fact, that evolutionary biologists would not regard a "synthetic dog" as a dog at all.

[255] Putnam, 1993, p.132.
[256] Putnam, 1993, p.131.

From their point of view, such a thing would simply be an artifact of no interest – unless, of course, it interbred either with natural dogs or with other synthetic dogs (and, in the latter case, became the start of the sort of "population" that evolutionary biologists study)."[257]

Again, we have the situation here where two explanations of the essence of doghood are constrained in the two ways I discussed earlier in this chapter. Both are concerned with regularities which are objectively out there in the real world. It is the case that dogs do have a particular evolutionary history. It is also the case that dogs have certain types of genetic structure. Furthermore, both explanations are suitably scientific. It seems clear that matters of descent, heredity and interbreeding will be important properties in determining and so giving an explanation of what it is to be a dog. Yet it seems that matters of genetic structure involving RNA and DNA will be equally important or relevant properties for explaining what it is to be a dog. To claim that one explanation is more basic, fundamental or scientific than the other would smack of arbitrariness. Rather, it seems that again we have two equally good, objective and so correct scientific explanations which concern themselves with different (scientific) characteristics and regularities. This illustrates that, even at the scientific level, natural kinds cannot be said to have one unique essence or explanation attached to them and which essence one takes to be definitive of the kind will change the way that counterfactual cases are decided.

Finally, I want to take a brief look at John Haldane's discussion of Putnam's molecular/evolutionary biologist case in his paper, "Humanism with a Realist Face". Haldane begins by asking what the principles of identity and individuation of dogs are. He claims, in reply, that, "...it is clear that the answers will have to account for the fact that dogs are living things born of living things" and that, "...any account of the nature of dogness must accommodate the fact that individuals of the species, and the species itself, are extended in time". He suggests that these requirements are at the root of the evolutionary biologist's historical account of essence. He goes on to claim, however, that such an account is, by itself, inadequate since, "...it gives no explanation of the source of the structure in the individual and, at the general level in the species, that contains and subserves those vital activities that are characteristic of dogs, explaining such things as their physiology and causal powers." He suggests that it is requirements such as these that inform the molecular biologist's account.

[257] Putnam, 1993, p.133.

The conclusion which Haldane draws from this is that an adequate and proper answer to the question 'what constitutes the essence of doghood?' must involve a combination of both the molecular and the evolutionary account. He says, "Putnam asks: which gives the essence of dogness – evolutionary or biological theory? I suggest the answer is "both", or rather, each gives an aspect of the answer. When those aspects are combined in ways that constitute a unified conception we shall have a yet more adequate concept of dogness and, ipso facto, the structure of the mind will be more fully that of the reality that informs it."[258] This is intended to show that Putnam has not succeeded in dismissing the Aristotelian notion that natural kinds have single determinate natures or essences, quite independent of human interests.

I cannot agree with Haldane that the evolutionary and molecular accounts are, by themselves inadequate, but if appropriately combined, give an adequate account of what it is to be a dog. It seems to me that Haldane is simply making a claim with no evidence to back it up. Indeed, we *could* argue that in order to give an adequate explanation of what it is to be a dog, we must give a combined molecular/ evolutionary account, but I do not think we are forced to do so. We could equally claim that the evolutionary and the molecular accounts each provide entirely adequate explanations of the essence of doghood and that which one you choose will depend on your interests, on what you find relevant or salient, as Putnam does. Despite Haldane's appeal to reality, I do not see that *reality* forces us to make one choice over and above the other. This is another instance of the general point which I have been arguing throughout this thesis – that reality is inherently flexible and so can support more than one classification and definition of the entities which make it up and that for this reason, we as human beings must make choices and decisions in order to get classification off the ground.

Therefore, to follow Putnam's strategy or to follow Haldane's strategy is not to be guided by reality, since reality does not provide us with guidelines here, rather it is to make a choice or decision concerning which of the several real properties or characteristics are to count for matters of classification. Haldane's argument simply provides us with a further illustration of my general thesis – that classification involves more than reflecting the way the world is and that epistemology plays an important role in the classificatory process.

Interestingly, Haldane fails to directly tackle the question posed by Putnam of

[258] All quotations come from Haldane, 1994, p.27.

whether the synthetic dog is or is not a dog. It seems, however, plausible to assume that he will want to claim that it is *not* a dog, since he is of the opinion that an adequate conception of what it is to be a dog involves a combination of the evolutionary and the molecular perspectives, yet this creature fails from the evolutionary perspective, since, to use Haldane's own terminology, it is not a "living thing" which has been "born of living things". Perhaps if this creature bred with a 'real' dog and produced 'real' puppies, Haldane might be tempted to claim that it was, after all, a dog. Yet even if this were the case, the creature still fails to fulfil Haldane's description completely, since it was not, itself, born of living things. At this stage, he would, presumably, need to explain why this particular aspect of the description is so important, given that this creature exhibits the same behaviour and molecular composition as a 'real' dog. It seems, then, that Haldane at least owes us some kind of explanation regarding what, exactly, this creature is and why it is such.

Finally, Haldane is clearly an objectivist. His claim that, "When those aspects are combined in ways that constitute a unified conception we shall have a yet more adequate concept of dogness and, ipso facto, the structure of the mind will be more fully that of the reality that informs it."[259] implies that there should be some kind of isomorphic correspondence between our concepts and the world and that there is some one ultimate description of that world which we ought to be aiming to reflect. The problems with this kind conception of the world and of our cognition of it have been discussed in detail in Chapter Two. Suffice it to repeat here that the world does not prefer Haldane's strategy over and above Putnam's and that each strategy is equally real, objective and correct. The spirit of this kind of approach is summed up well by Putnam in the context of observing that talk about the nature of things is relative to interests in much the same way that talk about causes is relative to the point of view of the person describing the cause. He says, "I remember discussing this some years ago with Richard Boyd who remarked that, while what counts as "the cause" of something may be relevant to interests, that something is the cause of something *given* those interests is absolute, and I am inclined to agree."[260] To translate: the world itself does not provide us with a unique classification of the entities within it, which means that we must approach that world with certain interests (underlying theories or explanations) which serve to make some of the similarities and regularities more salient than others. However, given a particular set

[259] Haldane, 1994, p.27.
[260] Putnam, 1993, p.134.

of interests or a particular theory or explanation, certain similarities and regularities will stand out as relevant and these similarities and regularities are real and objective features of the world. Putnam concludes, "Certainly, the Aristotelian insight that objects have structure is right, provided we remember that what counts as the structure of something is relative to the ways in which we interact with it."[261]

Conclusions.

In this chapter, we have seen that the psychological explanation-based account of categorisation serves as a suitable model of expert as well as of lay classification. This is because, even at the scientific level, the world can be divided in more than one way and the choice of one division over another invokes certain underlying theories or explanations.

Secondly, we have seen that, although it can be correct to claim that possession of certain essential properties explains regularity of behaviour and characteristics, there are different regularities running through the natural world which can be explained by different essential properties. Even operating with a narrow conception of what is to count as a scientific property, it simply does not seem to be the case that there is one explanation which is clearly the most basic or fundamental. What we have uncovered is a situation where there are different, but equally important bits of scientific explanation going on. The different bits of explanation simply concentrate on different aspects of the natural environment.

It is because of the very fact that the environment comprises these different aspects that Objectivism makes no sense, that the explanation-based account works as a model of classification by scientific experts and that it is possible to come up with competing yet equally real or objective divisions of that environment.

[261] Putnam, 1993, p.134.

Chapter 9
Conclusions.

In Chapter One, I outlined the two major aims of this thesis. These were:

- To illustrate that scientific classification comprises a combination of metaphysics and epistemology.
- And in view of this combination, to show that the psychological explanation-based account of categorisation is a suitable model of expert scientific as well as of lay classification.

Throughout the course of this thesis both of these aims have, I think, been achieved.

I began by presenting the particular contexts in which traditional views of classification and my own view of classification come to life. We saw that the doctrine of Objectivism involves a commitment to an external world which is structured in terms of entities and classes of entities, quite independently of human cognition or theorising and we saw that objectivist epistemology requires a pure reflection of that structure. Human beings therefore have no active role to play with respect to either metaphysics or epistemology. The classical view of categorisation involves reflecting the classes which are inherent and ready-defined in nature.

In opposition to Objectivism, I presented Putnam's internal realism and Johnson's experiential realism. We saw that the God's Eye view of the world which is required by Objectivism is impossible to attain and so meaningless. Rather, the world is always mediated via our experience and this experience involves prior conceptualisation and understanding. I argued that this kind of theoretical backdrop fits well with the requirements of my own account of classification – that classification is not simply a matter of reflecting metaphysics, but involves a degree of human input in the form of choices and decisions (epistemology) which is vital in order to get classification off the ground. We saw that all of this is consonant with the idea that categorisation is based upon real properties of the entities which are being categorised. Putnam and Johnson continually stress that we are in constant

interaction with the world of which we form a part.

Next, I gave a brief account of different psychological views of categorisation from the classical view up to the present day. What we saw here was that in the early days of psychological research, the doctrine of Objectivism was accepted without question, hence the assumption of the classical view of categorisation. Yet, as time moved on, many psychologists shifted from this very tightly defined, static view of classification to a much more flexible and dynamic position which began to suggest that classification is a more human-oriented affair than was previously supposed and that epistemology actually makes some active contribution to the process of classification. The discussion of George Rey's response to earlier work in the psychology of categorisation was important for two reasons. Firstly, Rey is the only philosopher I know of who has commented directly upon this work and secondly, his views presented us with a concrete example of an objectivist philosopher denying the claim which I have been making – that epistemology has an important role to play in the process of categorisation.

It was the case studies and philosophical arguments presented in Chapters Six and Seven which provided the bulk of the evidence in support of my two major claims – that categorisation comprises a mixture of metaphysics and epistemology and that the psychological explanation-based account of categorisation is a suitable model of expert scientific classification. From examining three case studies in the history and sociology of science (as supported by material from contemporary biology), we were able to extract a series of six interrelated points. These were that:

- Scientific classification is not a straightforward, simple or clear-cut affair.
- Scientific classification involves much more than charting the way the world is.
- Scientific classification involves human input in the form of aims, theories and commitments.
- The line between metaphysics and epistemology is blurred.
- There are problems with applying the objectivist doctrine of essentialism to scientific classification.
- Scientific case studies lend support to my hypothesis that the explanation-

based model of categorisation applies as much to classification by scientific experts as to classification by laypeople.

These six points really represent a more detailed breakdown of the two major claims which I make in this thesis and they therefore serve to combat traditional views of scientific classification, as revived by Putnam and Kripke, which embrace all the assumptions of the doctrine of Objectivism. By analysing the two papers by Keith Donnellan and John Canfield, we were able to build upon the evidence gleaned from these case studies. In particular, we produced more theoretical support for the claims that categorisation is more than the way the world is, that categorisation involves a degree of human input and that the line between metaphysics and epistemology is blurred.

Finally, we saw that my views on categorisation do not entail that the scientist can divide up the natural world in any way that he likes. We saw that the scientist is constrained by the limitation of utility and by the 'suitably scientific' limitation and that these two are linked by the fundamental notion of explanation. I went on to argue, however, that there is *not* one ultimately basic set of scientific explanations of the natural world which can be captured by providing a unique division of that world into kinds on the basis of possession of essential properties. I therefore prevented the possibility of an objectivist account of classification re-emerging at the level of scientific explanation.

To conclude, in this thesis I have shown, by means of a mixture of theoretical argumentation and empirical evidence, that scientific classification comprises a combination of metaphysics and epistemology and therefore involves the scientist in both discovery *and* decision-making.

Bibliography.

Ackrill, J.L. (ed.) (1987). *A New Aristotle Reader*. Oxford: Clarendon Press.

Aristotle (1963). *Categories and De Interpretatione*. (Trans. J.L. Ackrill). Oxford: Clarendon Press.

Aristotle (1972). *De Partibus Animalium I and De Generatione Animalium I*. (Trans. D.M. Balme). Oxford: Clarendon Press.

Barsalou, L.W. (1989). Intraconcept Similarity and its Implications for Interconcept Similarity. In S. Vosniadou and A. Ortony (eds.) *Similarity and Analogical Reasoning*. (pp. 76–121). Cambridge: Cambridge University Press.

Bellezza, F.S. (1984). Reliability of Retrieval from Semantic Memory: Noun Meanings. *Bulletin of the Psychonomic Society*, **22**, 377–380.

Bellezza, F.S. (1984). Reliability of Retrieval from Semantic Memory: Common Categories. *Bulletin of the Psychonomic Society*, **22**, 324–326.

Bloor, D. (1982). Durkheim and Mauss Revisited: Classification and the Sociology of Knowledge. *Studies in History and Philosophy of Science*, **13**, 267–297.

Canfield, J.V. (1983). Discovering Essence. In C. Ginet and S. Shoemaker (eds.) *Knowledge and Mind: Philosophical Essays*. (pp. 105–129). Oxford: Oxford University Press.

Churchland, P.S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

Cohen, J.M. and Cohen, M.J. (1971/1980). *The Penguin Dictionary of Modern Quotations*. (Second Edition). Harmondsworth: Penguin.

Cracraft, J. (1992). Species Concepts and Speciation Analysis. In M. Ereshefsky (ed.) *The Units of Evolution*. Cambridge, MA: MIT Press.

Davidson, D. (1967). Truth and Meaning. *Synthèse*, **17**, 304–323.

De Queiroz, K. (1992). Phylogenetic Definitions and Taxonomic Philosophy. *Biology and Philosophy*, **7**, 295–313.

Dean, J. (1979). Controversy Over Classification: A Case Study from the History of Botany. In B. Barnes and S. Shapin (eds.) *Natural Order*. (pp. 211–230). London: Sage Publications.

Desmond, A.J. (1979). Designing the Dinosaur: Richard Owen's Response to Robert Edmond Grant. *Isis*, **70**, 224–234.

Donnellan, K.S. (1983). Kripke and Putnam on Natural Kind Terms. In C. Ginet and S. Shoemaker (eds.) *Knowledge and Mind: Philosophical Essays*. (pp. 84–104). Oxford: Oxford University Press.

Dowty, D.R., Wall, R.E. and Peters, S. (1981). *Introduction to Montague Semantics*. Dordrecht, Holland: D. Reidel.

Dupré, J. (1981). Natural Kinds and Biological Taxa. *Philosophical Review*, **90**, 66–90.

Dupré, J. (1993). *The Disorder of Things*. Cambridge, MA: Harvard University Press.

Eddings, D. (1993). *The Elenium*. London: Harper Collins.

Ereshefsky, M. (1992). *The Units of Evolution*. Cambridge, MA: MIT Press

Field, H. (1982). Realism and Relativism. *Journal of Philosophy*, **79**, 553–567.

Fleck, L. (1979). *Genesis and Development of a Scientific Fact.* (Eds. T.J. Trenn and R.K. Merton). Chicago: University of Chicago Press.

Frege, G. (1952). On Sense and Meaning. In P. Geach and M. Black (eds.) *Translations from the Philosophical Writings of Gottlob Frege*. (Third edition). (pp. 56–78). Oxford: Basil Blackwell.

Goodman, N. (1972). Seven Strictures on Similarity. In N. Goodman, *Problems and Projects*. (pp. 24–32). Indianapolis and New York: Bobbs-Merrill.

Haldane, J. (1994). Humanism with a Realist Face. *Philosophical Books*, **35**, 21–29.

Hanson, N.R. (1958). *Patterns of Discovery*. Cambridge: Cambridge University Press.

Hughes, S.S. (1977). *The Virus: A History of the Concep*t. London: Heinemann Educational.

Hull, D.L. (1984). A Matter of Individuality. In E. Sober (ed.) *Conceptual Issues in Evolutionary Biology*. (pp. 623–645). Cambridge, MA: MIT Press.

Hull, D.L. (1988). *Science as a Process*. Chicago: University of Chicago Press.

Hull, D.L. (1989). *The Metaphysics of Evolution*. Albany: SUNY Press.

Johnson, M. (1994). *The Body in the Mind*. Chicago: University of Chicago Press.

Keil, F.C. (1989). *Concepts, Kinds and Cognitive Development*. Cambridge, MA:

MIT Press.

Kitcher, P. (1984). Species. *Philosophy of Science*, **51**, 308–333.

Komatsu, L.K. (1992). Recent Views of Conceptual Structure. *Psychological Bulletin*, **112**, 500–526.

Kripke, S.A. (1972). Naming and Necessity. In D. Davidson and G. Harman (eds.) *Semantics of Natural Language*. (pp. 253–355). Dordrecht, Holland: D. Reidel Publishing Company.

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lewis, D. (1972). General Semantics. In D. Davidson and G. Harman (eds.) *Semantics of Natural Language*. (pp. 169–218). Dordrecht, Holland: D. Reidel Publishing Company.

Locke, J. (1690/1964). *An Essay Concerning Human Understanding*. (Ed. A.D. Woozley). Glasgow: Collins.

Malt, B.C. (1990). Features and Beliefs in the Mental Representation of Categories. *Journal of Memory and Language*, **29**, 289–315.

Malt, B.C. (1994). Water is not H_2O. *Cognitive Psychology*, **27**, 41–70.

Malt, B.C. and Smith, E.E.. (1982). The Role of Familiarity in Determining Typicality. *Memory and Cognition*, **10**, 69–75.

Mayr, E. (1984). Species Concepts and their Application. In E. Sober (ed.) *Conceptual Issues in Evolutionary Biology*. (pp. 531–540). Cambridge, MA: MIT Press.

McCloskey, M.E. and Glucksberg, S. (1978). Natural Categories: Well Defined or Fuzzy Sets? *Memory and Cognition*, **6**, 462–472.

McNamara, T.P. and Sternberg, R.J. (1983). Mental Models of Word Meaning. *Journal of Verbal Learning and Verbal Behaviour*, **22**, 449–474.

Medin, D.L. (1989). Concepts and Conceptual Structure. *American Psychologist*, **44**, 1469–1481.

Medin, D. and Ortony, A. (1989). Psychological Essentialism. In S. Vosniadou and A. Ortony (eds.) *Similarity and Analogical Reasoning*. Cambridge: Cambridge University Press.

Mellor, D.H. (1977). Natural Kinds. *British Journal for the Philosophy of Science,* **28**, 299–312.

Mervis, C.B. and Pani, J.R. (1980). Acquisition of Basic Object Categories. *Cognitive Psychology*, **12**, 496–522.

Mishler, B.D. and Donoghue, M.J. (1992). Species Concepts: A Case for Pluralism. In M. Ereshefsky (ed.) *The Units of Evolution.* (pp.121–137) Cambridge, MA: MIT Press.

Murphy, G.L. and Medin, D.L. (1985). The Role of Theories in Conceptual Structure. *Psychological Review*, **92**, 289–314.

Putnam, H. (1975). The Meaning of "Meaning". In H. Putnam, *Mind, Language and Reality*. Cambridge: Cambridge University Press.

Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

Putnam, H. (1990). *Realism with a Human Face.* Cambridge, MA: Harvard University Press.

Putnam, H. (1993). Aristotle after Wittgenstein. In R.W. Sharples (ed.) *Modern Thinkers and Ancient Thinkers.* (pp. 117–137). London: UCL Press.

Putnam, H. (1994). Sense, Nonsense and the Senses: an Inquiry into the Powers of the Human Mind. *Journal of Philosophy*, **91**, 445–517.

Rey, G. (1983). Concepts and Stereotypes. *Cognition*, **15**, 237–262.

Rey, G. (1985). Concepts and Conceptions: A Reply to Smith, Medin and Rips. *Cognition*, **19**, 297–303.

Rips, L.J. (1989). Similarity, Typicality and Categorisation. In S. Vosniadou and A. Ortony (eds.) *Similarity and Analogical Reasoning*. Cambridge: Cambridge University Press.

Rosch, E. (1973). On the Internal Structure of Perceptual and Semantic Categories. In T.E. Moore (ed.) *Cognitive Development and the Acquisition of Language.* (pp. 111–145). San Diego, CA: Academic Press.

Rosch, E. (1978). Principles of Categorization. In E. Rosch and B. LLoyd (eds.) *Cognition and Categorization.* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E. and Mervis, C.B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, **7**, 573–605.

Smith, E.E. (1989). Three Distinctions about Concepts and Categorization. *Mind and Language*, **4**, 57–61.

Smith, E.E. and Medin, D.L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Smith, E.E., Medin, D.L. and Rips, L.J. (1984). A Psychological Approach to Concepts: Comments on Rey's "Concepts and Stereotypes". *Cognition*, **17**, 265–274.

Smith, E.E., Osherson, D.N., Rips, L.J. and Keane, M. (1984). *A Theory of Conceptual Combination for Prototype Concepts*. Unpublished manuscript, Massachusetts Institute of Technology.

Sober, E. (1993). *Philosophy of Biology*. Oxford: Oxford University Press.

Sokal, R.R. and Crovello, T.J. (1992). The Biological Species Concept: A Critical Evaluation. In M. Ereshefsky (ed.) *The Units of Evolution*. (pp. 27–55) Cambridge, MA: MIT Press.

Staniland, H. (1972). *Universals*. Garden City, NY: Doubleday.

Winsor, M.P. (1969). Barnacle Larvae in the Nineteenth Century: A Case Study in Taxonomic Theory. *Journal of the History of Medicine*, **24**, 294–309.

Wittgenstein, L. (1961). *Tractatus Logico Philosophicus*. (Trans. D.F. Pears and B.F. McGuinness). London: Routledge and Kegan Paul.

Zemach, E.M. (1976). Putnam's Theory on the Reference of Substance Terms. *Journal of Philosophy*, **73**, 116–127.